Preliminary Examination of Commercial Crewmember License Data

CFEC Report No. 07-7N
December, 2007

Prepared by Cathy Tide

Alaska Commercial Fisheries Entry Commission
8800 Glacier Highway, Suite 109
P.O. Box 110302
Juneau, Alaska 99811-0302
(907) 789-6160

OEO/ADA Compliance Statement

The Commercial Fisheries Entry Commission is administratively attached to the Alaska Department of Fish and Game (ADF&G).

ADF&G administers all programs and activities free from discrimination based on race, color, national origin, age, sex, religion, marital status, pregnancy, parenthood, or disability. The department administers all programs and activities in compliance with Title VI of the Civil Rights Act of 1964, Section 504 of the Rehabilitation Act of 1973, Title II of the Americans with Disabilities Act of 1990, the Age Discrimination Act of 1975, and Title IX of the Education Amendments of 1972.
If you believe you have been discriminated against in any program, activity, or facility please write:

- ADF&G ADA Coordinator, P.O. Box 115526, Juneau, AK 99811-5526
- U.S. Fish and Wildlife Service, 4401 N. Fairfax Drive, MS 2042, Arlington, VA 22203
- Office of Equal Opportunity, U.S. Department of the Interior, 1849 C Street NW MS 5230, Washington DC 20240.

The department's ADA Coordinator can be reached via phone at the following numbers:

(VOICE) 907-465-6077
(Statewide Telecommunication Device for the Deaf) 1-800-478-3648
(Juneau TDD) 907-465-3646
(FAX) 907-465-6078

For information on alternative formats and questions on this publication, please contact the following:

Commercial Fisheries Entry Commission (CFEC)
Research Section
8800 Glacier Highway, Suite 109
P.O. Box 110302
Juneau, Alaska 99811-0302
(907) 789-6160 *phone*
(907) 789-6170 *fax*

Abstract

The licensing system utilized by Alaska Department of Fish and Game for issuing commercial crewmember licenses is pragmatic; it enables participants in the commercial fisheries of Alaska to obtain licenses easily and across the state, including in remote locations. The licenses are issued annually with no information linking a license holder from one year to licenses held in other years. Despite this, the Commercial Fisheries Entry Commission is interested in tracking individual crew license holders through time and matching license holders to Commercial Fisheries Entry Commission permitting information when possible. Because commercial crewmember license data were not intended for use in this way, the data were evaluated to determine if it might be possible. This report summarizes the commercial crewmember license data for 1988 through 2006 and provides a brief summary on the quality of the data. The focus of this paper is on data fields which could be used to identify unique individuals or to define the residency of the license holder.

1.0 Introduction

Interest in commercial crewmember data has been growing in the public, the State, and with nongovernmental agencies. Groups are interested in studying commercial fishing employment trends, tracking crewmember participation in particular fisheries, and in light of potential allocation issues, developing histories for crewmembers based on participation in fisheries. A primary source of information about commercial crewmembers arises from commercial crewmember licenses issued by the Alaska Department of Fish and Game (ADF&G). According to ADF&G regulations, a person is required to obtain a commercial crewmember license in order to participate in commercial fishing in any waters of Alaska, if they do not already hold a valid Commercial Fisheries Entry Commission (CFEC) interim-use or limited entry permit card (5 AAC 39.110).[1] ADF&G has been issuing commercial crewmember licenses since 1988.[2] With the issuance of each license, certain data are collected about the license holder.

ADF&G Division of Administrative Services (DAS) and roughly 1,600 vendors across the state of Alaska sell hunting, trapping, sport fishing, and commercial crewmember licenses.[3] The sale of commercial crewmember licenses through vendors, rather than strictly ADF&G DAS, makes them readily accessible to persons interested in working in the fishing industry.[4] A crewmember can be licensed and able to work on short notice and in remote locations.

This licensing system was intended to make obtaining a crewmember license easy and quick for participants in the fishing industry. It was not intended to be a definitive source of demographic data on crew, or be a means for tracking crewmember participation. However, there is an interest in using crewmember data for this purpose. This raises the question whether the data can be successfully used in a way that was never intended.

CFEC, as well as other groups, may have similar questions about commercial crewmembers and crewmember license data. Even though not intended for this purpose, could the crew data be used for demographic analysis? Could license holders be followed through time? Would it be possible to identify individuals that have held crew licenses and also held interim-use or limited entry commercial fishing permits? The initial step towards answering these questions was to analyze the contents of ADF&G commercial crewmember data, and do a preliminary evaluation of its quality. Once those questions are answered, CFEC hopes to determine if individuals can be tracked through time or linked to CFEC permitting data. This report summarizes the commercial crewmember license data for 1988 through 2006 and provides a brief summary on the quality of the data. The focus of the paper is on data fields which could be used to identify unique individuals or define the residency of the license holder.

Key findings identified through the examination of the crew license data include:

- Fields in the crew license data vary from license year to license year. Response rates may vary between fields in each year and within a field for different years.
- Data submitted on licenses are captured as closely as possible to what is written on the license even if it appears incorrect. Fields may contain blank responses, filler data, and accurate as well as inaccurate responses.
- Roughly one-third of the license records between 1988 and 2006 do not contain information in the social security number field. It would be difficult to use the social security number as a unique identifier for individuals in the commercial crewmember data.
- Several fields in the commercial crewmember data could be used to estimate the residency of the license holder. Of those fields, the license class code may be the best to use since it has a high response rate in all years of license data and very few responses that fail to signify a residency.

[1] Commercial crewmember licenses enable crewmembers to participate in commercial fishing and qualify for benefits under the Alaska Fisherman's Fund. The Alaska Fishermen's Fund was established in 1951 to provide for the treatment of licensed commercial fishermen who were injured while fishing in Alaska.

[2] Commercial crewmember licenses were issued by the Alaska Department of Revenue between 1984 and 1987. According to an internal CFEC memo dated August 9, 2004, by Laura Joralemon, crewmember license data from 1984 through 1987 exist on floppy discs that may or may not be readable as data from the old WANG system that the Department of Revenue used. There are microfiche copies of the license data in report format. Paper copies of the licenses do not exist.

[3] Wright, Kristin. 2007. Personal communication. Alaska Department of Fish and Game; P.O. Box 115526, Juneau, AK 99811-5526.

[4] While there are roughly 1,600 vendors each year, vendor codes in the crew license data for non-voided licenses attribute only 236 to 352 vendors each year with selling commercial crewmember licenses between 1988 and 2006.

2.0 Crewmember Data

ADF&G DAS is currently responsible for issuing commercial crewmember licenses and distributing crewmember data to interested parties. Commercial crew licenses are sold through roughly 1,600 vendors across the state of Alaska as well as directly through ADF&G DAS. Individuals who purchase a commercial crewmember license through one of these vendors simply fill out a form to purchase the license. A copy of the form is retained by the crewmember, which acts as the license, and a copy of the license is submitted to ADF&G DAS by the vendor.

Each month, vendors report the number of licenses sold, by the type of license, to ADF&G DAS. That sale information is maintained by ADF&G DAS in a collection report database. The database contains counts of the number of crewmember licenses sold, by class code.[5] ADF&G DAS provides summary information on commercial crewmember license sales on their website (http://www.admin.adfg.state.ak.us/admin/license/licstats.html). The information posted includes the number of licenses sold by type, the overall number of resident and nonresident licenses sold, and a time series of license sales across years.

After vendors report the number of licenses sold in a month, they send copies of the actual licenses to ADF&G DAS. ADF&G DAS staff enter information from the license into the license file database, a second database containing crew license data. Information from each license is data captured as closely as possible to the written information on the license.[6] In recent years, licenses were also available from ADF&G DAS on the internet. That information is contained in the license file database as well. The license file database contains a list of all licenses issued by ADF&G DAS, including sport fishing, hunting, and trapping licenses. Commercial crewmember licenses are identified with a license type of 'BB' whereas hunting, trapping, and sport fishing licenses are designated with a license type of 'AA.'

The information in the license file database is maintained separately from the collection report file and the two are not synchronized. License information received from vendors after a cut-off date are not reflected in the collection report file. However vendors, for various reasons, send in prior year's licenses for up to several years. The license file reflects all the licenses sold, regardless of when vendors submitted the license information.[7] CFEC obtained commercial crewmember data from the license file database for license years 1988 through 2006. Since the license counts found on the ADF&G web site reflect the collection report file and will not match license counts found in the license file database, and the numbers presented in this report are derived from the license file database, the license counts presented herein will not match license counts found on the ADF&G web site.

The licensing system employed by ADF&G is successful in its goal to make commercial crewmember licenses readily available for those interested in participating in the commercial fishing industry. The relatively simple licensing system that is required to make this possible, and which contributes to the program's success, in part contributes to the drawbacks associated with using the existing commercial crewmember data for analysis. Licenses are issued annually. As such, there is no information linking a license holder from one year to licenses held in preceding years. Each crewmember has a different license number each year. Information about the license applicant is handwritten on licenses purchased through vendors. ADF&G DAS staff are required to interpret that handwriting when they enter the license information into the crew license database. The need to decipher handwriting can result in unintended errors in the data. Despite measures in place to prevent mistakes, data entry is always a source of potential error in data, as well. And finally, when someone purchases a license, ADF&G DAS staff must enter all of that person's information, even if it was entered earlier in the year for another license, or for licenses in prior years. This is another potential source for variability in license data.

[5] Class codes differentiate resident licenses from nonresident licenses, yearlong licenses from 7-day licenses, adult licenses from child licenses, and duplicate licenses from originally issued licenses.
[6] Even if information on the license appears incorrect (e.g.,, Anchorage, AR instead of Anchorage, AK) the data is captured as closely as possible to what is written on the license for enforcement purposes. Wright, Kristin. 2007. Personal communication. Alaska Department of Fish and Game; P.O. Box 115526, Juneau, AK 99811-5526.
[7] Wright, Kristin, personal communication.

2.1 Available Fields

A total of 61 data fields exist on commercial crewmember license data between 1988 and 2006. The fields contain information describing the license holder, the license, and data entry of license information. The fields found in the data vary from license year to license year. In general, earlier years contain fewer fields and more recent years have more fields. For example, there are 27 fields in the 1993 data and 54 fields in the 2005 data. In no years are all 61 data fields used. There are 4 fields that do not contain information in any year.[8] Table 1 indicates the 61 data fields found across all years of crewmember data and the years in which each field contains information. An 'X' indicates that at least one record in that license year contains an entry in the field.

2.2 Field Completeness

While it is important to note which fields contain entries in which years, it is equally important to know what percentage of records within a year contain an entry in that field. Table 2 indicates the percentage of records for a license year that contain an entry in each field. Within each license year it is clear that not all fields were collected and/or recorded equally. Some fields show a high percentage of response while others show less. For example, in 1988, 0.3% of the records contain an entry in the country field and 97.7% of the records contain an entry in the zip field. Within many fields, the percentage of records with entries varies from year to year. In some years there is a high response in a field and in other years the response is low. For example, in 1994, 0.8% of the records contain an entry in the country field, yet in 1995, the following year, 99.9% of the records contain an entry in that field.

It is important to note, however, that some records should not have an entry for particular fields. For example, an original license should have a value in the class code field, but should not provide a value for the original class code field. Conversely, a duplicate license should contain values in both the class code and original class code fields.[9] Only duplicate licenses should have responses in the original class code field. As such, the response rate for the original class code field should be much lower than the response rate for the class code field.

Although Table 2 indicates the percentage of records with a response for each field in a license year, it in no way evaluates the validity of the response. Records with an entry in a field may contain correct information, filler data, or inaccurate responses. Filler information was entered for some records when the licensee did not supply a value, but not always and not consistently between years or fields. In some cases, a '?' was entered if names were indecipherable, 01/01/1901 is commonly used as a filler value in the birth date field, and 000-00-0000, 001-01-0001, and 999-99-9999, are all used as filler data in the SSN field. Names, birth dates, and SSNs are also left blank when no information was provided by the licensee. As mentioned earlier, values in the commercial crewmember license data may be incorrect. Incorrect values can originate from both the license holder and be introduced during data entry of the license information.

The inconsistencies of data collection and errors in the data may make it difficult to draw conclusions about crewmember licenses across time.

[8] The 4 fields that do not contain a value in any record in any year are: MF2, RESULT_CODE, WC_ITEM_NUM, and WC_PERMIT_YR.
[9] Original licenses should indicate the type of license in the class code field, e.g., resident crew license, non-resident child license, 7-day resident license. Duplicate licenses, issued if the original is lost, should indicate a duplicate license in the class code field, and the type of license originally issued in the original class code field, e.g., resident crew license.

Table 1. The Presence of Data in a Data Field by License Year, For Crewmember License Data, 1988-2006.[1]

Field	License Year																		
LIC_YR	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006
LIC_NUM	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
TYPE_CODE	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
CLASS_CODE	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
ORIG_CLASS_CODE	X						X	X	X	X	X	X	X	X	X	X	X	X	X
FIRST_NAME	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
M_INITIAL	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
LAST_NAME	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
SUFFIX	X	X	X	X	X								X	X	X				
BIRTH_DATE	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
SSN	X	X	X	X	X	X	X				X	X	X	X	X	X	X	X	X
ADDRESS	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
CITY	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
STATE	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
ZIP	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
COUNTRY	X	X	X	X	X		X	X	X	X	X	X	X	X	X	X	X	X	X
SEX	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
RES_STATUS											X	X	X	X	X	X	X	X	X
RESIDENT_YEARS						X	X	X	X	X	X	X	X	X	X	X	X	X	X
RESIDENT_MONTHS	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
CITIZEN	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
DRIVER_STATE								X	X	X	X	X		X			X	X	
DRIVER_NUM								X	X	X	X	X		X			X	X	
WEIGHT	X	X	X								X	X	X	X	X	X	X	X	X
COLOR_EYES	X	X	X								X	X	X	X	X	X	X	X	X
COLOR_HAIR	X	X	X								X	X	X	X	X	X	X	X	X
HEIGHT_FT											X	X	X	X	X	X	X	X	X
HEIGHT_IN	X	X	X	X	X						X	X	X	X	X	X	X	X	X
MAIL_ADDRESS							X	X	X	X	X	X	X	X	X	X	X	X	X
MAIL_CITY							X	X	X	X	X	X	X	X	X	X	X	X	X
MAIL_STATE							X	X	X	X	X	X	X	X	X	X	X	X	X
MAIL_ZIP							X	X	X	X	X	X	X	X	X	X	X	X	X
MAIL_COUNTRY																		X	X
PHYSICAL_ADDRESS												X	X	X	X	X	X	X	X
EMAIL												X	X	X	X	X	X	X	X
PHONE_NUM															X	X	X	X	X
VENDOR	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
ISSUE_DATE	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
EFFECTIVE_DATE											X	X	X	X	X	X	X	X	X
BEG_DATE	X	X	X	X	X													X	X
END_DATE	X	X	X	X	X													X	X
ADDED_DATE	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X

Table 1. The Presence of Data in a Data Field by License Year, For Crewmember License Data, 1988-2006.[1]

Field	License Year																		
LIC_YR	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006
ADDED_USER	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
UPDATE_DATE	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
UPDATE_USER						X	X	X	X	X	X	X	X	X	X	X	X	X	X
BATCH_NUM	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
BATCH_SEQ_NUM	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
MF1	X	X	X	X	X														
MF2																			
DOC_DATE						X	X	X	X	X	X	X	X	X	X	X	X	X	X
DOC_NUM						X	X	X	X	X	X	X	X	X	X	X	X	X	X
RESULT_CODE																			
SOURCE											X	X	X	X	X	X	X	X	X
WC_ITEM_NUM																			
WC_PERMIT_YR																			
WEB_AUTHORIZATION												X	X	X	X	X	X	X	X
ORDER_NUM														X	X	X	X	X	X
GIFT_PURCH																		X	X
SHIP_ALT_ADDRESS																		X	X
PRINT_HOME_DATE																		X	X
CONFIRMATION_CODE																		X	X

[1] An ‘X’ indicates the field contains at least one record with an entry in the license year.

Table 2. The Percent of Records in a License Year with Data in a Field, For Crewmember License Data, 1988-2006.

Field	License Year																		
LIC_YR	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006
LIC_NUM	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
TYPE_CODE	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
CLASS_CODE	100.0%	100.0%	100.0%	100.0%	100.0%	99.9%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
ORIG_CLASS_CODE	0.1%						<0.1%	2.2%	2.0%	2.5%	2.8%	2.8%	2.6%	2.0%	1.7%	2.3%	2.3%	1.6%	2.2%
FIRST_NAME	96.7%	97.1%	97.3%	97.1%	97.2%	97.0%	97.3%	96.8%	97.0%	97.4%	96.8%	97.3%	95.8%	96.3%	96.4%	96.5%	96.3%	96.4%	96.6%
M_INITIAL	69.1%	69.8%	71.3%	74.1%	76.2%	74.3%	74.5%	80.3%	79.3%	80.4%	85.5%	83.0%	81.7%	84.0%	82.9%	83.6%	84.4%	83.4%	84.2%
LAST_NAME	99.9%	100.0%	100.0%	100.0%	100.0%	97.1%	97.4%	96.8%	97.0%	97.4%	96.8%	97.3%	95.8%	96.3%	96.4%	96.5%	96.3%	96.4%	96.5%
SUFFIX	4.7%	4.5%	4.2%	4.5%	4.6%								<0.1%	<0.1%	<0.1%				
BIRTH_DATE	100.0%	100.0%	100.0%	100.0%	100.0%	93.0%	92.9%	95.8%	95.9%	95.8%	95.3%	96.8%	95.2%	95.7%	95.8%	95.8%	95.8%	95.7%	95.8%
SSN	74.6%	77.5%	78.1%	76.4%	72.3%	66.4%	66.6%				93.3%	95.2%	93.1%	31.1%	91.7%	93.8%	93.7%	91.6%	93.9%
ADDRESS	94.6%	95.3%	96.4%	96.8%	96.9%	96.4%	96.9%	96.6%	96.8%	97.2%	96.7%	97.2%	95.7%	96.3%	96.3%	96.5%	96.3%	96.3%	96.5%
CITY	96.7%	96.8%	96.9%	97.1%	99.5%	96.7%	96.9%	96.7%	97.0%	97.4%	96.6%	97.2%	95.7%	96.2%	96.4%	96.5%	96.3%	96.3%	96.5%
STATE	96.5%	96.7%	96.7%	96.8%	99.4%	96.5%	96.8%	96.4%	96.7%	97.0%	96.2%	96.9%	95.4%	96.0%	96.2%	96.3%	96.1%	96.1%	96.3%
ZIP	97.7%	96.6%	96.5%	96.8%	99.3%	96.3%	96.8%	96.1%	96.4%	96.7%	95.9%	96.5%	95.0%	95.7%	95.7%	95.8%	95.5%	95.6%	95.5%
COUNTRY	0.3%	0.5%	0.6%	0.9%	1.2%		0.8%	99.9%	99.4%	99.8%	99.9%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
SEX	95.7%	95.8%	96.0%	95.9%	95.8%	95.5%	95.5%	96.1%	96.7%	96.8%	96.5%	97.3%	95.8%	96.3%	96.4%	96.5%	96.3%	96.3%	96.5%
RES_STATUS											0.4%	0.2%	0.3%	0.3%	0.5%	0.7%	1.4%	1.6%	2.6%
RESIDENT_YEARS						52.4%	52.0%	50.5%	54.5%	51.1%	51.5%	52.6%	47.9%	48.4%	51.8%	66.8%	59.2%	57.1%	52.7%
RESIDENT_MONTHS	84.5%	100.0%	100.0%	100.0%	100.0%	52.4%	51.7%	25.8%	33.7%	25.7%	23.9%	32.5%	26.3%	28.3%	38.4%	58.2%	53.9%	47.1%	45.3%
CITIZEN	96.0%	96.0%	95.8%	96.2%	96.0%	96.0%	96.3%	94.3%	96.2%	96.6%	96.4%	97.3%	95.8%	96.3%	96.4%	96.5%	96.3%	96.3%	96.5%
DRIVER_STATE								61.0%	62.6%	64.1%	65.0%	2.3%							
DRIVER_NUM								59.2%	66.6%	66.2%	67.6%	2.3%		<0.1%			<0.1%	<0.1%	
WEIGHT	98.6%	100.0%	99.9%								0.4%	0.2%	0.3%	0.3%	0.5%	0.7%	1.4%	1.6%	2.6%
COLOR_EYES	96.6%	96.7%	96.9%								0.4%	0.2%	0.3%	0.3%	0.5%	0.7%	1.4%	1.6%	2.6%
COLOR_HAIR	96.5%	96.4%	96.8%								0.4%	0.2%	0.3%	0.3%	0.5%	0.7%	1.4%	1.6%	2.6%
HEIGHT_FT											0.3%	0.2%	0.3%	0.3%	0.5%	0.7%	1.4%	1.6%	2.6%
HEIGHT_IN	98.5%	100.0%	100.0%	100.0%	100.0%						0.3%	0.2%	0.3%	0.3%	0.5%	0.7%	1.4%	1.6%	2.6%
MAIL_ADDRESS							96.9%	96.5%	96.8%	97.2%	96.7%	97.2%	95.7%	96.2%	96.3%	96.4%	96.3%	96.2%	96.5%
MAIL_CITY							96.9%	96.7%	97.0%	97.4%	96.6%	97.2%	95.7%	96.2%	96.4%	96.4%	96.3%	96.3%	96.5%
MAIL_STATE							96.8%	96.4%	96.7%	97.0%	96.2%	96.9%	95.4%	96.0%	96.1%	96.2%	96.1%	96.1%	96.3%
MAIL_ZIP							96.8%	96.1%	96.4%	96.7%	95.9%	96.5%	95.0%	95.6%	95.7%	95.7%	95.5%	95.5%	95.5%
MAIL_COUNTRY																		1.4%	2.5%
PHYSICAL_ADDRESS												<0.1%	0.1%	0.2%	0.4%	0.6%	1.2%	1.4%	2.4%
EMAIL												<0.1%	0.1%	0.1%	0.3%	0.4%	0.7%	0.8%	1.5%
PHONE_NUM															0.2%	0.3%	0.5%	0.7%	1.0%
VENDOR	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
ISSUE_DATE	100.0%	100.0%	100.0%	100.0%	100.0%	99.4%	99.7%	96.5%	96.7%	97.0%	96.1%	97.2%	95.7%	96.3%	96.2%	96.3%	96.1%	96.1%	96.3%
EFFECTIVE_DATE											0.4%	0.2%	0.3%	0.3%	0.5%	0.7%	1.3%	1.5%	2.5%
BEG_DATE	100.0%	100.0%	100.0%	100.0%	100.0%													1.4%	7.0%
END_DATE	100.0%	100.0%	100.0%	100.0%	100.0%													1.5%	7.0%
ADDED_DATE	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
ADDED_USER	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
UPDATE_DATE	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
UPDATE_USER						100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
BATCH_NUM	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	99.9%	99.8%	99.6%	99.4%	98.7%	98.6%	97.5%
BATCH_SEQ_NUM	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	99.9%	99.8%	99.6%	99.4%	98.7%	98.6%	97.5%
MF1	100.0%	100.0%	100.0%	100.0%	100.0%														

Table 2. The Percent of Records in a License Year with Data in a Field, For Crewmember License Data, 1988-2006.

Field	License Year																		
LIC_YR	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006
MF2																			
DOC_DATE						100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
DOC_NUM						100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
RESULT_CODE																			
SOURCE											0.4%	0.2%	0.3%	0.3%	0.5%	0.7%	1.4%	1.6%	2.6%
WC_ITEM_NUM																			
WC_PERMIT_YR																			
WEB_AUTHORIZATION												<0.1%	0.1%	0.2%	0.4%	0.6%	1.3%	1.4%	2.5%
ORDER_NUM														0.2%	0.4%	0.6%	1.3%	1.4%	2.5%
GIFT_PURCH																		1.4%	2.5%
SHIP_ALT_ADDRESS																		1.4%	2.5%
PRINT_HOME_DATE																		<0.1%	0.1%
CONFIRMATION_CODE																		<0.1%	0.1%

2.3 Valid Licenses

Commercial crewmember data contain licenses that may be valid or voided crewmember licenses. A voided license may be one that the vendor or licensee made a mistake on, subsequently voided, then filled out another license correctly, somewhat like voiding a check. A voided license may also reflect when the license fee payment is refunded to the licensee and ADF&G DAS staff void the license record to reflect the refund.[10] Voided licenses appear to be identified by several different methods across the years (and in some cases within the same year). Voided licenses have been identified with a class type of 'VD' or 'BV' and/or the last name of 'VOID', 'BV' or 'AV'. It is not clear if the different reasons for voided licenses are reflected by the different methods of indicating a voided license. Table 3 indicates the number of license observations in the crewmember data, the number that appear valid, and the number that appear to be voided, each by license year.

Table 3. Total, Valid, and Voided Crewmember Licenses with Method Indicating Void, 1988-2006.[1]

License Year	Total Licenses	Valid Licenses		Voided Licenses		Void Indication				
						Class Code		Last Name		
						VD	BV	AV	BV	VOID
1988	32,781	31,733	96.8%	1,048	3.2%					1,048
1989	33,725	32,750	97.1%	975	2.9%	747				975
1990	37,592	36,583	97.3%	1,009	2.7%					1,009
1991	37,759	36,738	97.3%	1,021	2.7%		24		24	997
1992	37,023	36,032	97.3%	991	2.7%		968	18	924	44
1993	34,042	33,056	97.1%	986	2.9%	975		3	1	7
1994	33,506	32,613	97.3%	893	2.7%	882		1	10	
1995	31,814	30,788	96.8%	1,026	3.2%	1,026				
1996	29,445	28,564	97.0%	881	3.0%	881				
1997	28,041	27,306	97.4%	735	2.6%	735				
1998	26,012	25,169	96.8%	843	3.2%	843				
1999	25,773	25,075	97.3%	698	2.7%	698				
2000	25,292	24,229	95.8%	1,063	4.2%	1,063				
2001	21,261	20,475	96.3%	786	3.7%	786				
2002	17,997	17,349	96.4%	648	3.6%	648				
2003	18,830	18,166	96.5%	664	3.5%	664				
2004	19,377	18,661	96.3%	716	3.7%	716				
2005	19,443	18,719	96.3%	724	3.7%	724				
2006	19,089	18,426	96.5%	663	3.5%	663				

[1] Some voided licenses are indicated by the class code, some by the last name field, and others with both fields. The number of void indicators will not necessarily sum to the number of voided licenses per license year.

The remainder of this document will refer to valid crewmember licenses.

3.0 Unique Identification

Since the primary objective of CFEC's examination of ADF&G commercial crewmember data is to explore whether individuals can be tracked through time or to see if crewmember data can be linked to other CFEC data, it is imperative that unique individuals in the license data be identified. Typically, an individual's Social Security Number is the primary piece of information used to identify unique individuals. Birth date and name are other fields that can be used to help identify unique individuals. The focus of the analysis on data availability and data quality will begin with these fields.

[10] George, Susan. 2007. Personal communication. Alaska Department of Fish and Game; P.O. Box 115526, Juneau, AK 99811-5526.

3.1 Social Security Numbers

Social Security Numbers (SSN) are commonly used as a unique identifier for an individual. The SSN field in commercial crewmember data contains responses for the vast majority of the crewmember licenses issued in a year, with the exception of 1995, 1996, 1997, and 2001. Since ADF&G DAS was told they could not collect SSN information at that time, no social security numbers were recorded for license years 1995, 1996, and 1997.[11] In 2001, a majority of license records do not have an entry in the social security number field. Table 4 indicates the number of valid licenses by license year that have entries in the SSN field and the number of licenses for which these data were not collected or recorded.

Although a response may appear in the SSN field, the information may or may not reflect a valid SSN value. According to the Social Security Administration (SSA), valid SSNs have not been issued with an area number (first 3 digits) greater than 772.[12] In addition, SSNs will never be issued with an area number of '000', a group number (middle 2 digits) of '00', or a serial number (final 4 digits) of '0000'. The SSN field in the commercial crew license data contains responses matching each of these invalid SSN situations. Table 4 indicates the number of entries in the SSN field, by license year, that do not represent legitimate SSN values based on criteria presented by the SSA, and the number of responses in the SSN field that do.

According to the Internal Revenue Service (IRS), an individual taxpayer identification number (ITIN) can be issued to someone who does not have, or is not eligible to obtain an SSN.[13] It is possible that crewmembers may indicate an ITIN on their license in place of an SSN. However, less than 0.1% of the responses in the SSN field in each license year reflect an ITIN format.[14] For the purposes of this report, any ITIN entered in the SSN field will be treated as an SSN response.

Even if the response in the SSN field has an acceptable SSN or ITIN format, it may or may not be the actual SSN or ITIN of the person obtaining the license. The SSN field may contain inaccurate information for many reasons. A few include: inadvertent errors on the part of the license applicant, like transposition of numbers; the intentional use of an incorrect value, to maintain confidentiality of one's SSN; misinterpretation of handwriting on the license by ADF&G staff; and data entry errors. In order to be as inclusive as possible, for the purposes of this report and subsequent analyses, only responses in the SSN field that are fewer than 9 digits, contain non-numeric values such as '.', or are ADF&G filler values (e.g., 999-99-9999, 000-00-0000, 001-01-0001), will be flagged as invalid. All others will be flagged as valid, even if they violate the criteria presented by the SSA which are described above. If a crewmember has selected a number to use as an identifier, in place of a true SSN, it is possibly used for a license in more than one year or for more than one license.

Table 4 indicates the number of licenses by license year that have responses in the SSN field which follow the format constraints for SSN values as defined by the SSA and those that do not. The table also indicates the number of licenses, by license year, for which the entry in the SSN field has been flagged as invalid or valid for purposes of this study.[15] Due to a heavier use of the ADF&G filler value 001-01-0001 between 1990 and 1994, a higher number of licenses have an SSN flagged as invalid as have an SSN that truly violates the correct format for SSN values in those years. This filler value is not found on licenses after the 1994 license year. Over all years, 1988 through 2006, 67.6% of the licenses contain SSN responses flagged as valid.[16] With nearly one third of the valid crew licenses lacking usable SSN entries, it would be difficult to use the SSN as a unique identifier for crewmembers.

Nearly all SSNs appear only once in the crewmember license data per license year, but some appear more than once. Between 1988 and 2006, SSNs appear up to as many as six times in a single license year. Table 5 indicates the number of unique social security numbers per license year, and the number of times unique SSNs appear on a crewmember license in a license year. Over all years, 1988 through 2006, 96.5% of the unique SSNs appear only once per year.

---

[11] Wright, Kristin, personal communication.
[12] http://www.ssa.gov/
[13] http://www.irs.gov/
[14] Individual taxpayer identification numbers are nine-digit numbers that begin with a 9 and have a 7 or 8 as the 4th digit (e.g., 9XX-7X-XXXX or 9XX-8X-XXXX).
[15] Social security numbers were flagged as invalid if they have fewer that 9 digits, contain non-numeric characters, or were an ADF&G filler value (e.g., 999-99-9999, 000-00-0000, or 001-01-0001).
[16] Excluding license years 1995, 1996, 1997, and 2001, 84.0% of licenses contain SSN responses flagged as valid.

Table 4. The Number of Valid Licenses with SSN Data, Valid SSN Formats, Invalid SSN Formats, SSN Values Flagged as Invalid, and SSN Values Flagged as Valid, for License Years 1988-2006.

License Year	Valid Licenses	License Contains No SSN Data		License Contains Some SSN Data		License with Invalid SSN Format[1]		License with Valid SSN Format[1]		SSN Value Flagged as Invalid[2]		SSN Value Flagged as Valid[2]	
1988	31,733	7,273	22.9%	24,460	77.1%	70	0.2%	24,390	76.9%	42	0.1%	24,418	77.0%
1989	32,750	6,600	20.2%	26,150	79.8%	76	0.2%	26,074	79.6%	67	0.2%	26,083	79.6%
1990	36,583	7,247	19.8%	29,336	80.2%	96	0.3%	29,240	79.9%	123	0.3%	29,213	79.9%
1991	36,738	7,880	21.5%	28,858	78.5%	75	0.2%	28,783	78.3%	108	0.3%	28,750	78.2%
1992	36,032	9,265	25.7%	26,767	74.3%	62	0.2%	26,705	74.1%	112	0.3%	26,655	74.0%
1993	33,056	10,457	31.6%	22,599	68.4%	68	0.2%	22,531	68.2%	89	0.3%	22,510	68.1%
1994	32,613	10,305	31.6%	22,308	68.4%	28	0.1%	22,280	68.3%	58	0.2%	22,250	68.2%
1995	30,788	30,788	100.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%
1996	28,564	28,564	100.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%
1997	27,306	27,306	100.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%
1998	25,169	911	3.6%	24,258	96.4%	24	0.1%	24,234	96.3%	3	< 0.1%	24,255	96.4%
1999	25,075	533	2.1%	24,542	97.9%	26	0.1%	24,516	97.8%	1	< 0.1%	24,541	97.9%
2000	24,229	686	2.8%	23,543	97.2%	23	0.1%	23,520	97.1%	0	0.0%	23,543	97.2%
2001	20,475	13,854	67.7%	6,621	32.3%	7	< 0.1%	6,614	32.3%	0	0.0%	6,621	32.3%
2002	17,349	853	4.9%	16,496	95.1%	11	0.1%	16,485	95.0%	0	0.0%	16,496	95.1%
2003	18,166	506	2.8%	17,660	97.2%	17	0.1%	17,643	97.1%	2	< 0.1%	17,658	97.2%
2004	18,661	503	2.7%	18,158	97.3%	23	0.1%	18,135	97.2%	3	< 0.1%	18,155	97.3%
2005	18,719	919	4.9%	17,800	95.1%	13	0.1%	17,787	95.0%	0	< 0.1%	17,800	95.1%
2006	18,426	501	2.7%	17,925	97.3%	22	0.1%	17,903	97.2%	3	< 0.1%	17,922	97.3%

[1] Social security numbers are nine-digit numbers. Values with less than 9 digits have an invalid format. SSNs with an area number (first 3 digits) of '000' or greater than '772', or a group number (middle 2 digits) of '00', or a serial number (final 4 digits) of '0000' also have an invalid format. Individual taxpayer identification numbers are nine-digit numbers that begin with a 9 and have a 7 or 8 as the 4th digit, e.g., 9XX-7X-XXXX or 9XX-8X-XXXX. Values in any other format are invalid ITINs. Percentage reflects the percent of valid licenses that have SSNs with either a valid or invalid format.

[2] Social security numbers are flagged as invalid if they contain fewer than 9 digits, contain non-numeric characters, or are ADF&G filler values (e.g., 000-00-0000, 001-01-0001, or 999-99-9999). All other SSNs are flagged as valid for the purposes of this study. Percentage reflects the percent of valid licenses that have SSNs which have been flagged as valid or invalid.

Table 5. Unique Social Security Numbers (SSNs) Flagged as Valid Per License Year and the Number of Times Each SSN Appears on a Crewmember License Per License Year, 1988-2006.

Liecnse Year	Unique SSNs	Number of Appearances on a Crewmember License											
		1		2		3		4		5		6	
1988	23,716	23,047	97.2%	639	2.7%	27	0.1%	3	<0.1%				
1989	25,097	24,168	96.3%	875	3.5%	51	0.2%	3	<0.1%				
1990	28,179	27,215	96.6%	903	3.2%	52	0.2%	9	<0.1%				
1991	27,701	26,726	96.5%	905	3.3%	67	0.2%	2	<0.1%	1	<0.1%		
1992	25,759	24,905	96.7%	815	3.2%	36	0.1%	3	<0.1%				
1993	21,766	21,064	96.8%	662	3.0%	38	0.2%	2	<0.1%				
1994	21,473	20,744	96.6%	685	3.2%	40	0.2%	4	<0.1%				
1995													
1996													
1997													
1998	23,333	22,467	96.3%	815	3.5%	46	0.2%	5	<0.1%				
1999	23,562	22,651	96.1%	846	3.6%	62	0.3%	3	<0.1%				
2000	22,696	21,912	96.6%	730	3.2%	45	0.2%	9	<0.1%				
2001	6,391	6,169	96.5%	214	3.4%	8	0.1%						
2002	15,990	15,515	97.0%	445	2.8%	29	0.2%	1	<0.1%				
2003	17,021	16,420	96.5%	565	3.3%	36	0.2%						
2004	17,520	16,938	96.7%	538	3.1%	37	0.2%	6	<0.1%			1	<0.1%
2005	17,206	16,657	96.8%	511	3.0%	34	0.2%	2	<0.1%	1	<0.1%	1	<0.1%
2006	16,980	16,244	95.7%	597	3.5%	82	0.5%	47	0.3%	10	<0.1%		

The appearance of SSNs on multiple licenses may be due to several reasons: an individual may obtain multiple 7-day licenses within the same year, each valid for a different 7 days; an individual may obtain duplicate licenses in the same year in the event the original license is lost; data entry errors may result in the same SSN being used for different individuals; or different individuals may obtain licenses in the same year, or different years for that matter, using the same social security number. The first two examples are valid reasons for an SSN to appear multiple times in a license year. The third reason is unfortunate but unavoidable when human error may occur during data entry or human judgment is required to decipher handwriting. Since there is no immediate validation or proof required of the SSN written when purchasing a license, it is possible that license holders may write the incorrect SSN or transpose numbers within the SSN. The final reason accounting for an SSN to appear more than once in a license year – more than one individual using the same SSN – further demonstrates why it is not possible to use the SSN reported on the commercial crewmember licenses as a unique identifier for individuals.

3.2 Birth Dates

Birth dates can serve as a secondary way to help identify or distinguish crewmembers. Crewmember birth dates have been recorded from licenses each year between 1988 and 2006. Table 6 indicates the number of valid licenses by license year that have entries in the birth date field and the number of licenses for which this data was not collected or recorded.[17] Of the licenses that contain birth date responses, the birth dates that appear to be possible are differentiated from those that are improbable. Birth dates are identified as improbable if they result in a crewmember being less than 5 years old or greater than 85 years old at the time of license issuance (or year end if the issue date is not available).[18] A higher percentage of licenses have birth dates that appear possible than SSNs that have been flagged as valid. When sorting the crew license data by either name or SSN, it becomes clear through visual inspection that birth date responses are also susceptible to data entry errors, handwriting interpretation, and possible mistakes by the license holder.

Table 6. The Number of Licenses with Birth Date Data, Birth Date Data Which Appear Possible, Improbable Birth Date Data, and No Birth Date Data, 1988-2006.

License Year	Valid Licenses	License Contains No Birth Date Data		License Contains Some Birth Date Data		Birth Date Data Appears Improbable[1]		Birth Date Data Appears Possible	
1988	31,733	11	<0.1%	31,722	>99.9%	1,827	5.8%	29,895	94.2%
1989	32,750	10	<0.1%	32,740	>99.9%	1,931	5.9%	30,809	94.1%
1990	36,583	5	<0.1%	36,578	>99.9%	2,123	5.8%	34,455	94.2%
1991	36,738	4	<0.1%	36,734	>99.9%	1,724	4.7%	35,010	95.3%
1992	36,032	2	<0.1%	36,030	>99.9%	1,696	4.7%	34,334	95.3%
1993	33,056	1,394	4.2%	31,662	95.8%	142	0.4%	31,520	95.4%
1994	32,613	1,503	4.6%	31,110	95.4%	121	0.4%	30,989	95.0%
1995	30,788	314	1.0%	30,474	99.0%	141	0.5%	30,333	98.5%
1996	28,564	319	1.1%	28,245	99.0%	116	0.4%	28,129	98.5%
1997	27,306	451	1.7%	26,855	98.3%	104	0.4%	26,751	98.0%
1998	25,169	372	1.5%	24,797	98.5%	82	0.3%	24,715	98.2%
1999	25,075	140	0.6%	24,935	99.4%	105	0.4%	24,830	99.0%
2000	24,229	158	0.7%	24,071	99.3%	86	0.4%	23,985	99.0%
2001	20,475	129	0.6%	20,346	99.4%	81	0.4%	20,265	99.0%
2002	17,349	107	0.6%	17,242	99.4%	92	0.5%	17,150	98.9%
2003	18,166	140	0.8%	18,026	99.2%	56	0.3%	17,970	98.9%
2004	18,661	102	0.6%	18,559	99.4%	48	0.3%	18,511	99.2%
2005	18,719	130	0.7%	18,589	99.3%	54	0.3%	18,535	99.0%
2006	18,426	145	0.8%	18,281	99.2%	38	0.2%	18,243	99.0%

[1] Birth dates are identified as improbable if they result in a crewmember being less than 5 years old or greater than 85 years old at the time of license issuance (or year end if the issue date is not available). Birth date filler data (e.g., 01-01-1901) are flagged as improbable.

[17] Birth dates that are filler data (e.g., 1901-01-01) are included as licenses with birth date data but included with those flagged as improbable.
[18] Age was calculated by determining the number of days between the birth date and license issue date (if available) or between the birth date and December 31 of the license year (if the issue date was unavailable) and then the days were converted to years.

3.3 Names

Names can also serve as a way to help identify or distinguish crewmembers. Crewmember license data contains fields for first, middle, and last names as well as name suffix (e.g., Jr., Sr., III). Table 7 indicates the number of valid licenses by license year that has any combination of first, middle initial and last name response. Nearly all the license data have first and last or first, middle initial and last name entries. Visual inspection of the data identifies that name data are also susceptible to data entry errors and handwriting interpretation. Difference in the name fields are also compounded by the use of full names, nicknames, initials, name changes (e.g., due to marriage), and abbreviations. For example, William, Bill, Billy, Wm. and W. may all be used for the same individual.

4.0 Residency

Residency is a characteristic that is often the focus of commercial fishing discussions and data analyses. Are vessel owners or permit holders residents or nonresidents? What percentage of commercial harvest is by Alaska residents? It is therefore likely that analysis of commercial crew member data would also entail an investigation of license holders by residency. There are several fields in the commercial crew license data that could be used to estimate the residency of the license holder. Those fields are examined here for availability and data quality.

4.1 License Class Code

The license class code field is used to distinguish different types of commercial crewmember licenses. The code can differentiate resident licenses from nonresident licenses, yearlong licenses from 7-day licenses, and adult licenses from child licenses. The class code can also distinguish duplicate licenses from originally issued licenses. The license class code field is present in each year of commercial crewmember data, and contains a value for 99.9 to 100.0% of the records for each license year. Because of the high response rate for the class code field, it appears to be a good candidate for estimating residency.

Class codes of '30', '34', and '36' indicate resident crew, resident child, and resident 7-day commercial crewmember licenses, respectively. Class codes of '31', '35', and '37' indicate nonresident crew, nonresident child, and nonresident 7-day commercial crewmember licenses, respectively. A vast majority of licenses indicate one of the resident or nonresident licenses. Table 8 indicates the number of valid commercial crewmember licenses that have resident or nonresident class codes in each license year.

There are a small percentage of valid licenses that indicate a class code which does not reflect the residency of the license.[19] Duplicate licenses, indicated by class code '32' or '39' do not reflect residency. When these licenses are purchased, the license holders are supposed to indicate the class code from the original license. Of the 18,945 duplicate licenses issued between 1988 and 2006, 65.8% did not indicate the class code from the original license (12,460 licenses). A resident license, either crew or child crew, was indicated for 4,164 licenses (22.0%) and a nonresident crew or nonresident child crew license was indicated for 2,290 duplicate licenses (12.1%). Reporting the original class code on duplicate license purchases improved dramatically in 1995, yet never beyond 79.9% of the duplicate licenses in a license year.

In 2000 and 2001 a child crew license was available for the first time.[20] This type of license did not specify resident or nonresident status. Resident and nonresident child crew licenses, indicated by class codes '34' and '35', respectively, were available beginning in 2002. The child crew licenses issued in 2000 and 2001, which do not indicate residency, only account for 1.7% of the licenses issued in those years, and 0.1% of licenses over all years between 1988 and 2006. If duplicate licenses are excluded, and since unclassified child licenses comprise such a small percentage of licenses issued, the class code field serves as a reasonable field for estimating the residency of the license holder.

[19] Between 2.9% and 5.3% of licenses each license year indicate a class code which does not signify residency.
[20] A child crew license is indicated by class code '33'.

Table 7. Name Information Available on Valid Crew License Data by License Year, 1988-2006.

				Name Data Available													
License Year	Valid Licenses	No Name Data		First		First & Middle		First & Last		First & Middle & Last		Middle & Last		Last			
1988	31,733	17	<0.1%	2	<0.1%	1	<0.1%	9,065	28.6%	22,634	71.3%	1	<0.1%	13	<0.1%		
1989	32,750							9,198	28.1%	23,532	71.9%	2	<0.1%	18	<0.1%		
1990	36,583							9,781	26.7%	26,785	73.2%	1	<0.1%	16	<0.1%		
1991	36,738							8,715	23.7%	27,965	76.1%	1	<0.1%	57	<0.1%		
1992	36,032							7,776	21.6%	28,222	78.3%	3	<0.1%	31	<0.1%		
1993	33,056							7,735	23.4%	25,278	76.5%			43	0.1%		
1994	32,613							7,634	23.4%	24,973	76.6%	1	<0.1%	5	<0.1%		
1995	30,788	1	<0.1%			1	<0.1%	5,247	17.0%	25,539	83.0%						
1996	28,564			3	<0.1%			5,203	18.2%	23,358	81.8%						
1997	27,306			2	<0.1%			4,772	17.5%	22,532	82.5%						
1998	25,169	2	<0.1%					2,936	11.7%	22,231	88.3%						
1999	25,075	1	<0.1%	1	<0.1%			3,682	14.7%	21,390	85.3%			1	<0.1%		
2000	24,229					1	<0.1%	3,562	14.7%	20,665	85.3%			1	<0.1%		
2001	20,475							2,613	12.8%	17,862	87.2%						
2002	17,349			1	<0.1%			2,435	14.0%	14,913	86.0%						
2003	18,166							2,426	13.3%	15,740	86.7%						
2004	18,661							2,315	12.4%	16,346	87.6%						
2005	18,719							2,529	13.5%	16,190	86.5%						
2006	18,426							2,366	12.8%	16,060	87.2%						

Table 8. License Count by License Class Code and License Year, 1988-2006.

License Year	Valid Licenses	Resident Crew Licenses					Nonresident Crew Licenses					Unclassified Crew Licenses[1]				Unknown License Types[2]	Non-Crew License Types[3]
		30 Crew	34 Child	36 7-day	Total		31 Crew	35 Child	37 7-day	Total		33 Child	32 or 39 Dup.	Total			
1988	31,733	19,559	0	0	19,559	61.6%	10,443	0	0	10,443	32.9%	0	1,287	1,287	4.1%	442	2
1989	32,750	20,139	0	0	20,139	61.5%	11,261	0	0	11,261	34.4%	0	1,349	1,349	4.1%	1	0
1990	36,583	21,237	0	0	21,237	58.1%	13,982	0	0	13,982	38.2%	0	1,353	1,353	3.7%	11	0
1991	36,738	20,602	0	0	20,602	56.1%	14,619	0	0	14,619	39.8%	0	1,497	1,497	4.1%	20	0
1992	36,032	20,064	0	0	20,064	55.7%	14,643	0	0	14,643	40.6%	0	1,324	1,324	3.7%	1	0
1993	33,056	18,420	0	0	18,420	55.7%	13,282	0	0	13,282	40.2%	0	1,317	1,317	4.0%	37	0
1994	32,613	18,174	0	0	18,174	55.7%	13,075	0	0	13,075	40.1%	0	1,364	1,364	4.2%	0	0
1995	30,788	16,696	0	0	16,696	54.2%	12,917	0	0	12,917	42.0%	0	1,175	1,175	3.8%	0	0
1996	28,564	15,643	0	0	15,643	54.8%	11,911	0	0	11,911	41.7%	0	1,010	1,010	3.5%	0	0
1997	27,306	14,634	0	0	14,634	53.6%	11,772	0	0	11,772	43.1%	0	900	900	3.3%	0	0
1998	25,169	13,404	0	0	13,404	53.3%	10,815	0	0	10,815	43.0%	0	950	950	3.8%	0	0
1999	25,075	13,801	0	0	13,801	55.0%	10,355	0	0	10,355	41.3%	0	918	918	3.7%	0	1
2000	24,229	12,673	0	0	12,673	52.3%	10,283	0	0	10,283	42.4%	408	865	1,273	5.3%	0	0
2001	20,475	10,489	0	0	10,489	51.2%	8,904	0	0	8,904	43.5%	355	725	1,080	5.3%	2	0
2002	17,349	9,213	302	0	9,515	54.8%	7,231	38	0	7,269	41.9%	0	563	563	3.2%	0	2
2003	18,166	9,623	296	0	9,919	54.6%	7,612	24	0	7,636	42.0%	0	610	610	3.4%	0	1
2004	18,661	9,753	308	0	10,061	53.9%	7,928	22	0	7,950	42.6%	0	650	650	3.5%	0	0
2005	18,719	9,809	328	102	10,239	54.7%	7,656	37	230	7,923	42.3%	0	556	556	3.0%	0	1
2006	18,426	8,958	313	295	9,566	51.9%	7,136	24	1,168	8,328	45.2%	0	532	532	2.9%	0	0

[1] Some license class codes do not specify residency and are called 'Unclassified' for the purposes of this report.
[2] Class codes of ' ', 'BB', '40', and '42' are indicated for some commercial crewmember licenses. These are not valid class codes, and licenses with this class codes are grouped as 'Unknown license types' for the purposes of this report.
[3] The information for these licenses indicates a commercial crewmember license (license type 'BB') but class codes inappropriate for a commercial crewmember license. Class code 18 refers to a duplicate sport fishing license and class code 3B refers to a resident trapping license.

4.2 Resident Status

Resident status is another field in the commercial crewmember license data which, based on its name, would appear useful in residency determination. This field originated in 1998, but unfortunately, is only used infrequently. The highest percentage of records with a response in this field occurs in 2006 and is 2.6%. With such a low response rate, this field should not be used in residency determination.

4.3 Resident Months

Data for the resident months field has been collected each year. Between 1988 and 1992 the field was heavily used with between 84.5 and 100.0% of the records containing an entry in the field, depending on the license year. Usage of the field decreased after 1992. Response for the resident months field was roughly 50.0% or lower by license year between 1993 and 2006. The decrease in response may be the result of the introduction of a resident years field in 1993.

The data in the resident months field has been recorded in a variety of formats over the years. There are responses collected as 4 digits using leading zeros (e.g., 0008), 2 digits using leading zeros (e.g., 08) and without any leading zeros (e.g., 8). To further complicate the matter, some responses are character values (e.g., 'FE' or 'NE'). Following a brief clean-up of the entries in the field, the numeric responses result in residency claims between 0 and 107.25 years.[21] Less than 0.01% of the commercial crewmember licenses result in residency claims 85 years or greater, however.

4.4 Resident Years

The resident years field was first recorded in 1993. Roughly 50% of the records each year between 1993 and 2006 contain responses in this field. The responses are recorded in a variety of formats over the years. The responses are collected as 2 digits with leading zeros (e.g., 06) and without any leading zeros (e.g., 6). Some of the responses were recorded as characters (e.g., 'LI' or 'UN').

Following a brief clean-up of the entries in the field, the numeric responses result in residency claims between 0 and 98 years.[22] Again, less than 0.01% of the commercial crewmember licenses result in residency claims 85 years or greater based on the year field alone.

4.5 Apparent Residency

An apparent residency time was estimated based upon the resident months and resident years fields combined. The resident months and resident years fields were converted to the same numeric format and all character responses were converted to zero. The resident months field was converted to a calculated years field by dividing the month response by 12 (e.g., 462 months divided by 12 results is 38.5 years). Adding the years and calculated years fields created a total residency field. License holders were flagged as apparent residents if their total residency field was greater than or equal to one year.[23] Table 9 indicates the number of licenses that have

[21] As a cursory clean-up of the resident month field, data formats were unified and character values were converted to 0.

[22] Just as with the resident month field, a cursory clean-up was performed on the resident years field in which data formats were unified and character values converted to 0.

[23] Alaska statutes (AS 16.05.415(a)) define a resident as a person (including an alien) who is physically present in Alaska with the intent to remain indefinitely and make a home here, has maintained that person's domicile in Alaska for the 12 consecutive months immediately preceding this application for a license, and is not claiming residency or obtaining benefits under a claim of residency in another state, territory, or country; a member of the military service or U.S. Coast Guard who has been stationed in Alaska for the 12 consecutive months immediately preceding this application for a license; or a dependent of a resident member of the military service or U.S. Coast Guard who has lived in Alaska for the 12 consecutive months immediately preceding this application for a license. A person who does not otherwise qualify as a resident may not qualify by virtue of an interest in an Alaska business. Per AS 16.05.415(b): A person who establishes residency in the state in accordance with the residency provision above remains a resident during an absence from the state unless during the absence the person (1) establishes or claims residency in another state, territory, or country; or (2) performs an act, or is absent under circumstances, that are inconsistent with the intent required under the residency provision above.

been issued to individuals that are estimated to be residents based on their responses to the resident months and/or resident years fields on the crewmember license.

Using this approach, license holders may be identified as nonresidents for several reasons: they indicated in the resident months and/or resident years field a length of time less than 1 year; a non-numeric response was indicated in the fields; no response was provided in these fields; a combination of a non-numeric and a blank value was used in the fields; and filler data was entered by ADF&G staff (e.g., 0, 00, 0000) when no data were provided by the license holder. Table 9 also indicates why nonresident license holders were flagged as such. Those licenses whose license holders indicated a length of time less than 1 year, including zero, and those licenses where an ADF&G staff member entered a numeric filler value are classified as nonresidents "due to numeric data." All others are classified as nonresidents "due to non-numeric or missing data."[24]

Between 1988 and 1992, license holders are predominately flagged as nonresidents based on actual numeric information. From 1993 on, however, nearly all of the license holders flagged as nonresidents are identified so because no information was available on the license or non-numeric responses were provided. Since missing data or non-numeric responses are responsible for such a high percentage of nonresident flags, the residency months and residency years, and as such the calculated apparent residency field, may not be reliable for estimating the residency of license holders.

Table 9. Apparent Residency by License Year and Reason for Nonresident Assignment, 1988-2006.[1]

License Year	Valid Licenses	Residents		Nonresidents		Nonresident			
						Due to Numeric Data[2]		Due to Non-Numeric or Missing Data[3]	
1988	31,733	19,913	62.8%	11,820	37.2%	7,148	60.5%	4,672	39.5%
1989	32,750	19,481	59.5%	13,269	40.5%	13,269	100.0%	0	0.0%
1990	36,583	20,288	55.5%	16,295	44.5%	16,295	100.0%	0	0.0%
1991	36,738	19,840	54.0%	16,898	46.0%	16,898	100.0%	0	0.0%
1992	36,032	19,550	54.3%	16,482	45.7%	16,482	100.0%	0	0.0%
1993	33,056	17,727	53.6%	15,329	46.4%	100	0.7%	15,229	99.3%
1994	32,613	17,318	53.1%	15,295	46.9%	107	0.7%	15,188	99.3%
1995	30,788	15,816	51.4%	14,972	48.6%	329	2.2%	14,643	97.8%
1996	28,564	14,819	51.9%	13,745	48.1%	378	2.8%	13,367	97.2%
1997	27,306	14,036	51.4%	13,270	48.6%	311	2.3%	12,959	97.7%
1998	25,169	13,025	51.8%	12,144	48.2%	515	4.2%	11,629	95.8%
1999	25,075	12,253	48.9%	12,822	51.1%	1,403	10.9%	11,419	89.1%
2000	24,229	11,639	48.0%	12,590	52.0%	578	4.6%	12,012	95.4%
2001	20,475	9,871	48.2%	10,604	51.8%	482	4.6%	10,122	95.4%
2002	17,349	8,515	49.1%	8,834	50.9%	855	9.7%	7,979	90.3%
2003	18,166	9,199	50.6%	8,967	49.4%	3,398	37.9%	5,569	62.1%
2004	18,661	9,441	50.6%	9,220	49.4%	2,049	22.2%	7,171	77.8%
2005	18,719	9,792	52.3%	8,927	47.7%	1,344	15.1%	7,583	84.9%
2006	18,426	9,297	50.5%	9,129	49.5%	787	8.6%	8,342	91.4%

[1] Apparent residency was calculated using responses to the resident months and resident years fields. Any license holder with a calculated residency of greater than or equal to 1 year was flagged as an apparent resident.
[2] The license holder was classified as a nonresident based on a numeric value in either the resident months or resident years field. The field resulted in residency less than 1 year.
[3] The license holder was classified as a nonresident because there was no numeric value in either the resident months field or the resident years field. The resident months and resident years fields were either blank, contained character values, or one of each.

[24] In some instances the character responses are ambiguous and a conversion to 0 seems appropriate (e.g., 'T4'). In others they appear to indicate nonresidence (e.g., 'NA' or 'NR'). In still others, it appears that the character responses in the residency months and residency years fields combined may actually indicate residency (e.g., 'LI' for months and 'FE' for years. Taken together, does this signify a resident for life?).

4.6 State

State is part of the address data collected for each commercial crewmember license. The responses in the state field contain more values than the 50 state abbreviations, US territories, and 13 Canadian provinces and territories combined. In addition to responses that do not appear to be states, territories, or provinces, the state field contains numeric and blank entries. A preliminary clean-up was performed on the state field.[25] Table 10 indicates the number of resident and nonresident licenses based on the corrected state field for each license year. All responses other than 'AK' are classified as nonresident.

Table 10. Resident and Nonresident Commercial Crewmember Licenses Based on a Corrected State, 1988-2006.

License Year	Valid Licenses	Resident Licenses		Nonresident Licenses	
1988	31,733	21,621	68.1%	10,112	31.9%
1989	32,750	21,981	67.1%	10,769	32.9%
1990	36,583	23,157	63.3%	13,426	36.7%
1991	36,738	22,596	61.5%	14,142	38.5%
1992	36,032	21,866	60.7%	14,166	39.3%
1993	33,056	20,141	60.9%	12,915	39.1%
1994	32,613	19,920	61.1%	12,693	38.9%
1995	30,788	18,197	59.1%	12,591	40.9%
1996	28,564	16,983	59.5%	11,581	40.5%
1997	27,306	15,841	58.0%	11,465	42.0%
1998	25,169	14,262	56.7%	10,907	43.3%
1999	25,075	14,917	59.5%	10,158	40.5%
2000	24,229	14,091	58.2%	10,138	41.8%
2001	20,475	11,713	57.2%	8,762	42.8%
2002	17,349	10,125	58.4%	7,224	41.6%
2003	18,166	10,586	58.3%	7,580	41.7%
2004	18,661	10,803	57.9%	7,858	42.1%
2005	18,719	10,954	58.5%	7,765	41.5%
2006	18,426	10,246	55.6%	8,180	44.4%

4.7 Citizenship

Data for the citizen field has been collected each year. A majority of responses are either Y or N and indicate whether the license holder is a U.S. citizen or not. In each year a small percentage of the licenses contain a response which does not indicate citizenship, such as blank values, question marks, and 2, F, M, T, and U. The 2, F, M, T, and U responses are only seen in 1988, however. Table 11 shows the citizenship indicated on crewmember license data for 1988 through 2006. A vast majority of license holders indicated that they are U.S. citizens.

[25] If the city indicated on a crewmember license was Anacortes, Bellingham, Beremertion (presumably Bremerton), Tukwilla, Tukwla (presumably Tukwilla), Port Townsend, Port Orchard, or Seattle, and the state indicated was 'AK', then the state field was corrected to 'WA.' If the city was Pebble Beach, Chico, or Daly City and the state was 'AK' then the state field was corrected to 'CA". If the city indicated on a license was Astoria and the state indicated was 'AK' then the state field was corrected to 'OR.' The state field was not modified for any other licenses.

Table 11. Citizenship of Commercial Crewmember License Holders, 1988-2006.

License Year	Total Valid Licenses	Blank		?		N		Y		2,F,M,T, or U	
1988	31,733	267	0.8%			362	1.1%	31,093	98.0%	11	<0.1%
1989	32,750	389	1.2%			562	1.7%	31,799	97.1%		
1990	36,583	575	1.6%			987	2.7%	35,021	95.7%		
1991	36,738	422	1.2%			912	2.5%	35,404	96.4%		
1992	36,032	502	1.4%			989	2.7%	34,541	95.9%		
1993	33,056	393	1.2%			804	2.4%	31,859	96.4%		
1994	32,613	353	1.1%			752	2.3%	31,508	96.6%		
1995	30,788	806	2.6%	36	0.1%	1,026	3.3%	28,920	93.9%		
1996	28,564	232	0.8%	90	0.3%	1,027	3.6%	27,215	95.3%		
1997	27,306	212	0.8%	565	2.1%	1,045	3.8%	25,484	93.3%		
1998	25,169	92	0.4%	777	3.1%	882	3.5%	23,418	93.0%		
1999	25,075	3	<0.1%	161	0.6%	963	3.8%	23,948	95.5%		
2000	24,229	3	<0.1%	190	0.8%	905	3.7%	23,131	95.5%		
2001	20,475	1	<0.1%	206	1.0%	767	3.8%	19,501	95.2%		
2002	17,349	1	<0.1%	111	0.6%	670	3.9%	16,567	95.5%		
2003	18,166			146	0.8%	604	3.3%	17,416	95.9%		
2004	18,661			124	0.7%	622	3.3%	17,915	96.0%		
2005	18,719	4	<0.1%	122	0.7%	518	2.8%	18,075	96.6%		
2006	18,426			143	0.8%	534	2.9%	17,749	96.3%		

4.8 Comparison of Residency Indicators

There are several fields in the commercial crewmember data that could be used to estimate the residency of the license holder. Table 12 provides a comparison of three: license class code, apparent residency, and state residency.[26] The table indicates the number of licenses issued between 1988 and 2006 by class code. The class codes are grouped by the residency associated with each, either resident, nonresident, unclassified, unknown, or non-crew. A subtotal of the number of licenses issued is provided for each grouping of class codes. The table then indicates for each class code how many of the licenses are defined as resident or nonresident by apparent residency and by state residency. For example, 284,835 licenses are considered resident licenses based on their class codes of 30, 34, and 36 (resident, child resident, and 7-day resident, respectively). Apparent residency flags 257,360 (90.4%) of those licenses holders as residents and 27,475 (9.6%) as nonresidents. State residency flags 282,239 (99.1%) of the license holders as residents and 2,596 (0.9%) as nonresidents.

This comparison clearly identifies the number of license holders that are estimated to be nonresidents, either by state address or length of residence, but who purchase resident licenses and conversely those estimated to be residents who purchase nonresident licenses. Apparent residency appears the most restrictive for allowing a resident classification (271,820 licenses across all license types) and state residency, based on the corrected state field, the most liberal (310,000 licenses across all license types). Class code falls between the two with 284,835 resident licenses between 1988 and 2006.

When class codes do not signify a residency, such as with unclassified, unknown, and non-crew class codes, apparent residency and state residency do indicate residency. There are 20,229 licenses with unclassified, unknown, or non-crew class codes; 11,999 (59.3%) are estimated as resident by apparent residency and 13,998 (69.2%) are estimated as resident according to state residency. The state and apparent residencies enable estimates of the residencies of license holders when the class code is unable to do so.

It will be important to consider which residency classification to use when proceeding with analysis of the commercial crewmember data. Class codes do not indicate residency for all licenses sold between 1988 and

[26] State address residency is based on a corrected state address field and apparent residency is based on the resident months and resident years fields.

2006. The numbers which do not indicate residency are a fairly small percentage, however. Apparent residency is considered the most conservative for estimating residency, since a high percentage of licenses are flagged as nonresident because the resident months and resident years fields contain no response or non-numeric responses. State residency is the most liberal in defining a resident license holder, but looks simply at the address of the license holder. It does not take into account the definition of resident set forth in the Alaska statutes (AS 16.05.415) and indicated on the license.

Table 12. Comparison of Resident or Nonresident License Classification by Class Code, Apparent Residency, and State Residency, for Commercial Crewmember Licenses, 1988-2006.

Class Code and License Type by Residency[1]	License Counts[2]	Apparent Residency[3]		State Residency[4]	
		Resident	Nonresident	Resident	Nonresident
Resident					
30 – Resident	282,891	255,644	27,247	280,317	2,574
34 – Child	1,547	1,359	188	1,538	9
36 – 7-day	397	357	40	384	13
Subtotal	284,835	257,360	27,475	282,239	2,596
Nonresident					
31 – Nonresident	205,825	2,422	203,403	13,661	192,164
35 – Child	145	6	139	20	125
37 – 7-day	1,398	33	1,365	82	1,316
Subtotal	207,368	2,461	204,907	13,763	193,605
Unclassified					
32 – Duplicate	18,940	11,129	7,811	12,950	5,990
33 – Child	763	548	215	664	99
39 – Duplicate	5	4	1	5	0
Subtotal	19,708	11,681	8,027	13,619	6,089
Unknown or Non-Crew					
Blank – Unknown	37	3	34	37	0
18 – Dup. Sport	5	2	3	4	1
40 – Unknown	1	0	1	0	1
42 – Unknown	1	1	0	1	0
BB – Unknown	475	310	165	335	140
3B – Res. Trapping	2	2	0	2	0
Subtotal	521	318	203	379	142

[1] The class codes indicated on commercial crewmember licenses (e.g., 30) and description of the class code (e.g., resident crewmember license), are grouped by associated residency. 'Unclassified' refers to crewmember license class codes which do not indicate a residency, 'Unknown' refers to class codes which do not exist, and 'Non-Crew' refers to class codes intended for licenses which are not commercial crewmember licenses.
[2] License counts reflect the number of commercial crewmember licenses issued based on class code.
[3] Apparent residency is based on a residence time calculated from the resident month and resident year fields from commercial crewmember licenses. A license holder with less than 1 year of residence time is considered nonresident and a license holder with 1 year or more is considered resident.
[4] State residency is based on the corrected state address field found on commercial crewmember licenses. Nonresidents have anything but 'AK' and residents have 'AK'.

5.0 Summary

The licensing system utilized by ADF&G for issuing commercial crewmember licenses is practical in that it enables participants in the commercial fisheries of Alaska to obtain a license quite easily, in populous as well as rural areas of the state. The information collected with the sale of each commercial crewmember license has become the focus of growing interest. Groups are interested in studying commercial fishing employment trends, tracking crewmember participation in particular fisheries, and developing a system through which crewmembers can build participation history.

CFEC is among the groups interested in commercial crewmember data. CFEC's interest in the data focuses on tracking an individual crew license holder through time and answering such questions as: how many years do persons typically work as a crewmember? or, how many commercial crew license holders have also held CFEC interim-use or limited entry permits? The data collected from the sale of licenses were never intended for this purpose. Licenses are issued annually and as such, a license holder from one year cannot necessarily be linked to licenses held in preceding years. Each crewmember has a different license number each year. It is CFEC's hope that the crew license data, although not intended for this purpose, can be used to track individuals through time.

The first step in evaluating whether this is possible was to review the existing data. Preliminary analysis revealed that there were inconsistencies in the data collection which may make it difficult to draw conclusions about crewmember licenses across time. Different data fields were collected in the 19 years of existing crew license data, and response rates vary between fields within a license year, and within a field across license years. In some cases, filler values are used when no information was provided by the license holder, and in others the field was left blank or contains erroneous entries.

Due to the nature of the crew license system, rigorous validation of data fields is not required before a crew license is issued to an individual. As such, the second purpose of this paper was to review the quality of the data available, with a focus on fields that could be used to identify unique individuals, or to estimate the residency of the license holder. Could the data available be used to answer policy questions about Alaska crewmembers?

Unique individuals might be identifiable using fields such as social security number (SSN), birth date, and name. Nearly one-third of the valid crew licenses issued between 1988 and 2006 lack usable information in the SSN field. That, combined with the occurrence of multiple people using the same SSN or errors in some SSN entries, makes it unlikely that SSN could be used as a unique identifier. Name and birth date information are available at a higher rate than SSNs in the data, yet these fields are susceptible to their own kinds of problems. Data contain improbable birth dates which result in crew license holders younger than 5 and older than 85 years of age. Name data for the same person can vary across observations as a result of full names, nicknames, initials, name changes, and abbreviations all being used.

Residency is commonly the focus of commercial fishing discussions. As such, the second focus of this paper was on the data which could infer the residency of the license holder. Several fields can be used to describe residency: license class code, residency status, resident months, resident years, and state address. Of those, class codes do not indicate residency for all licenses sold between 1988 and 2006, but if duplicate licenses are excluded, only a small percentage remains without a residency indicator. This appears the most reliable data in defining resident and nonresident license holders. Residency status is provided in very few of the license data, resident months and resident years contain a relatively high percentage of blank or non-numeric values, and the state address does not take into account the residency requirement set forth in the Alaska Statutes.

The next step in attempting to track an individual through time, and answering questions about crewmember participation, is to try to identify unique individuals found in the crew license data and assign unique identification numbers.