

Analysis of 2006 Stub Data

Judith Bjorndahl

March 26, 2019

Contents

| | |
|---------------------------------|----------|
| Data Representation | 1 |
| As an excel file | 1 |
| In R | 3 |
| Recommendations | 6 |
| Data storage | 6 |
| Survey design | 7 |
| Flexibility of survey | 8 |
| Analysis | 9 |
| By Survey | 9 |
| By Species | 9 |
| By County | 9 |
| By County and Species | 10 |

This report is an analysis of self-reported data from recreational anglers from 2006. It is an example of how the data from 2006 to 2016 can be processed and analyzed in R, with the ultimate goal of being able to quantify trends in the data over time. The first section is an explanation of the data as it is provided in an excel file, followed by an explanation of how it is transformed into a workable dataframe in R. The next section contains recommendations for how data can be collected and stored in future surveys to make the analyzing process simpler and more efficient (fewer survey responses discarded when calculating final statistics). Finally, the report also contains several different examples of how the data can be summarized for analysis.

Data Representation

The data are recorded from license stubs which recreational anglers fill out with information on the number of fish caught for ten different species in each NS county and the number of days spent fishing for each species. They also have the option to indicate that they did not fish or did not catch anything that season.

As an excel file

The data are stored in the file "2006 Stub Data.xls", which contains data from 3080 angler surveys (license stubs). There are 215 columns (variables). The first two columns give information on the survey responder:

- the first column is the license number (an integer with 2 to 5 digits)

2016 RESIDENT GENERAL FISHING LICENCE REPORT CARD Licence No. **16-040510**

Please record the number of days fished and the number of fish caught including released fish in each county and return the completed report card following your last fishing trip of the season.

FOR A CHANCE TO WIN VALUABLE PRIZES, PLEASE RETURN THIS COMPLETED CARD AFTER YOUR LAST DAY FISHED. GO TO NOVASCOTIA.CA/FISH/SPORTFISHING FOR DETAILS.

| | ANNAPOLIS | ANTIGONISH | CAPE BRETON | COLCHESTER | CUMBERLAND | DIGBY | GUYSBOROUGH | HALFAX | HANTS | INVERNESS | KINGS | LUNenburg | PICTOU | QUEENS | RICHMOND | SHELburne | VICTORIA | YARMOUTH | NO. OF DAYS FISHED |
|-------------------|-----------|------------|-------------|------------|------------|-------|-------------|--------|-------|-----------|-------|-----------|--------|--------|----------|-----------|----------|----------|--------------------|
| SPECKLED TROUT | | | | | | | | | | | | | | 4 | | | | | |
| BROWN TROUT | | | | | | | | | | | | | | | | | | | |
| RAINBOW TROUT | | | | | | | | | | | | | | | | | | | |
| SMALLMOUTH BASS | | | | | | | | | | | | | | | | | | | |
| WHITE PERCH | | | | | | | | | | | | | | | | | | | |
| YELLOW PERCH | | | | | | | | | | | | | | | | | | | |
| STRIPED BASS | | | | | | | | | | | | | | | | | | | |
| SHAD | | | | | | | | | | | | | | | | | | | |
| CHAIN PICKEREL | | | | | | | | | | | | | | | | | | | |
| LANDLOCKED SALMON | | | | | | | | | | | | | | | | | | | |

CHECK HERE IF NO FISH CAUGHT
 CHECK HERE IF YOU DID NOT FISH

Figure 1: A license stub from 2016.

| | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R | S | T | U | V | W | X | Y | Z | AA | AB | AC | AD | | | | | | | | | | | | | | | |
|-----|-------------------------|--------|----------|-------------|------------|----|-----------------------|--------|--------|------|------|----|-----|-----|-------|-------|------|-------|-------|-------|-------|--------|--------|------|------|-----|-----|------------|------|--------|--------|------|------|--|--|--|--|--|--|--|--|--|--|--|--|
| 1 | 2006 Stub Return | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 2 | | | | | | | SPECKLED TROUT | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 3 | | | | | | | Days | #count | County | ANNA | ANTI | CB | COL | CUM | DIGBY | GUY'S | HALL | HANTS | INVER | KINGS | LUNEN | PICTOU | QUEENS | RICH | SHEL | VIC | YAR | Effort Co. | Days | #count | County | ANNA | ANTI | | | | | | | | | | | | |
| 4 | Lic. Num | Res/Wk | No catch | Didn't fish | Effort Co. | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 574 | 421 | R | | | x | 2 | 1 | | | | 5 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 575 | 420 | R | | | x | 2 | 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 576 | 430 | R | | | x | 1 | 1 | | 5 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 577 | 441 | 1 day | | | x | 1 | 1 | | | | | | 3 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 578 | 441 | R | | | x | 1 | 1 | | 5 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 579 | 442 | R | | | x | 3 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 580 | 445 | R | | | x | 3 | 1 | | 10 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 581 | 448 | R | | | x | 5 | 1 | | 10 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 582 | 456 | R | | | x | 2 | 1 | | 6 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 583 | 461 | R | | | x | 6 | 1 | | 10 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 584 | 467 | R | | | x | 21 | 2 | | 27 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 585 | 431 | 1 day | | | x | 1 | 1 | | | | | | 1 | | | | | | | | 2 | | | | | | | | | | | | | | | | | | | | | | | | |
| 586 | 507 | R | | x | | | 1 | | 0 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 587 | 507 | N | | | x | 35 | 1 | | | | | | | | | | | | 27 | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 588 | 508 | R | | x | | | | | 0 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 589 | 508 | N | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 590 | 510 | R | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 591 | 521 | R | | | | 6 | 0 | | 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 592 | 527 | N | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 593 | 529 | R | | | x | 14 | 1 | | 52 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 594 | 532 | R | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 595 | 530 | R | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 596 | 540 | N | | | x | 5 | 1 | | | | | | 5 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 597 | 574 | R | | | x | 4 | 1 | | 6 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 598 | 579 | R | | | x | 15 | 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 599 | 591 | N | | | x | 15 | 1 | | | | | | 48 | | 15 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 600 | 606 | R | | x | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 601 | 654 | N | | | x | 4 | 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 602 | 656 | N | | | x | 4 | 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 603 | 663 | R | | x | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 604 | 675 | R | | | x | 4 | 1 | | 15 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 605 | 691 | R | | | x | 3 | 1 | | 15 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 606 | 682 | R | | x | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 607 | 684 | 1 day | | | x | 1 | 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Figure 2: A screenshot of the file "2006 Stub Data.xls" showing the first section of columns (columns E-Y are associated with Speckled Trout).

- the second column denotes whether the license holder is a resident (“R”), non-resident (“N”), or a 1 day license holder (“1 day”)

The next two columns are the responses of each survey respondent to the “no fish caught” and “did not fish” options. A checked box is indicated with “x” and an unchecked box is indicated with “” (empty string).

The remaining columns are organized into ten sections of 21 columns each, one section per species. Within each section, the first three columns are related to calculating catch per unit effort (CPUE), and the remaining 18 columns list the number of fish caught in each county (as an integer). Of the effort-related columns, one (“Days”) is raw data collected from the stubs and the other two are derived from the data:

- “# count.” is the number of counties in which the survey respondent fished for **the given species** (an integer)
- “Effort Calc” indicates whether the associated entry can be used to calculate CPUE for the species/county combination (“x”) or not (“”)
 - usually the requirements to be used for CPUE calculations are that the effort information (Days) is given and the respondent only fished for the associated species in one county
 - only some observations can be used for county CPUE calculations because the effort information is given on a per-species basis; if an angler fishes for the same species in multiple counties, it is impossible to tell how many days were spent in each county from the given data
 - this logic is not consistently applied across the board (these inconsistencies are addressed further in the recommendation section of the report)
 - some of the “Effort Calc” columns have integer values

In addition to the problems with integer values in the “Effort Calc” columns, there are some non-integer values and values of “0” in some of the “Days” columns.

In R

Several changes are made to how the data are stored in R to make processing simpler:

- the most major change in how the data are stored is the conversion from wide data format to long data format
 - each unique reported catch (county/species combination) becomes its own observation
 - there are new variables representing the county and species of the catch and the number of fish caught becomes its own, separate variable
- the responses to the “no fish caught” and “did not fish” questions are changed to logical variables (TRUE or FALSE, with “x” indicating TRUE)
- the values in the “Effort Calc” columns that are not “x” or “” are changed to NA values
- the values in the “Days” columns that are not integers or have a value of 0 are changed to NA values
- because some license numbers appear in the data more than once, a new variable, survey number, is created that is unique to each entry in the original data

Table 1: Some of the entries where the license numbers are not unique.

| Survey No. | License No. | Residency | Fished (Reported) | Caught (Reported) |
|------------|-------------|-----------|-------------------|-------------------|
| 713 | 1986 | R | TRUE | TRUE |
| 714 | 1986 | 1 day | TRUE | FALSE |
| 715 | 1988 | R | TRUE | TRUE |

| Survey No. | License No. | Residency | Fished (Reported) | Caught (Reported) |
|------------|-------------|-----------|-------------------|-------------------|
| 716 | 1988 | 1 day | TRUE | FALSE |
| 1163 | 11052 | R | TRUE | TRUE |
| 1164 | 11052 | N | TRUE | TRUE |
| 1172 | 11187 | N | TRUE | TRUE |
| 1173 | 11187 | R | TRUE | FALSE |
| 1789 | 25581 | R | TRUE | FALSE |
| 1790 | 25581 | R | TRUE | TRUE |
| 3005 | 57905 | R | TRUE | TRUE |
| 3006 | 57905 | R | TRUE | TRUE |

- three other new variables are derived from the data
 - a new logical variable indicating whether or not any fish were caught to accompany the “no fish caught” and “did not fish” survey responses
 - the number of counties in which a respondent fished for a given species is re-calculated in R
 - the “Effort Calc” variable is recalculated based on the rule that data can be used for county CPUE calculations if the data includes the number of days fished and the respondent only fished for the given species in one county
 - the differences in these (re)calculated variables and the corresponding pre-existing variables is discussed in the next two subsections

Problems with survey responses

The two check boxes on the survey form allow respondents to indicate that they did not fish or did not catch anything while fishing. However, some individuals either reported not fishing or not catching anything while also reporting a number of fish caught:

Table 2: Two anglers reported not fishing but caught fish.

| Lic. No. | Fished (Reported) | Caught (Rep.) | Caught (Actual) | Species | County | No. Fish |
|----------|-------------------|---------------|-----------------|----------------|--------|----------|
| 11625 | FALSE | NA | TRUE | Speckled Trout | GUYS | 8 |
| 11625 | FALSE | NA | TRUE | Speckled Trout | HALI | 81 |
| 11625 | FALSE | NA | TRUE | Speckled Trout | PICTOU | 12 |
| 47868 | FALSE | NA | TRUE | Speckled Trout | GUYS | 2 |

Table 3: Four anglers reported not catching anything but caught fish.

| Lic. No. | Fished (Reported) | Caught (Rep.) | Caught (Actual) | Species | County | No. Fish |
|----------|-------------------|---------------|-----------------|-----------------|--------|----------|
| 23291 | TRUE | FALSE | TRUE | Speckled Trout | LUNEN | 8 |
| 50521 | TRUE | FALSE | TRUE | Speckled Trout | ANTI | 40 |
| 50521 | TRUE | FALSE | TRUE | Speckled Trout | HALI | 4 |
| 50521 | TRUE | FALSE | TRUE | Speckled Trout | YAR | 2 |
| 50521 | TRUE | FALSE | TRUE | Brown Trout | ANTI | 3 |
| 50521 | TRUE | FALSE | TRUE | Smallmouth Bass | HALI | 450 |
| 50521 | TRUE | FALSE | TRUE | Chain Pickerel | HALI | 1 |
| 51356 | TRUE | FALSE | TRUE | Speckled Trout | SHEL | 10 |
| 61438 | TRUE | FALSE | TRUE | Speckled Trout | GUYS | 20 |
| 61438 | TRUE | FALSE | TRUE | Speckled Trout | HALI | 10 |

| Lic. No. | Fished (Reported) | Caught (Rep.) | Caught (Actual) | Species | County | No. Fish |
|----------|-------------------|---------------|-----------------|-------------|--------|----------|
| 61438 | TRUE | FALSE | TRUE | Brown Trout | GUYS | 1 |

Additionally, some anglers did not report any numbers of fish caught but did not indicate that they did not fish or that they did not catch anything.

Table 4: Four anglers did not report the number of fish caught.

| Lic. No. | Fished (Reported) | Caught (Rep.) | Caught (Actual) | Species | County | No. Fish |
|----------|-------------------|---------------|-----------------|----------------|--------|----------|
| 1546 | TRUE | TRUE | FALSE | NA | NA | NA |
| 22581 | TRUE | TRUE | FALSE | Speckled Trout | GUYS | 0 |
| 37174 | TRUE | TRUE | FALSE | Speckled Trout | NA | NA |
| 53879 | TRUE | TRUE | FALSE | NA | NA | NA |

While the first two inconsistencies are impossible, this inconsistency is possible if individuals are choosing to withhold information or did not remember how many fish were caught. However, it appears that the individual who reported “0” fish may have genuinely not caught anything. Additionally, there appears to be a data entry problem with one of these individuals, where the number of fish caught may have been entered in the wrong columns:

Table 5: Numerical values in the “Effort Calc” columns of raw data for one survey respondent.

| Lic. Num | Res/Non | Effort Calc. (SMB) | Effort Calc. (Strp. Bass) | Effort Calc. (Shad) |
|----------|---------|--------------------|---------------------------|---------------------|
| 1546 | R | 30 | 14 | 15 |

The inconsistencies in reporting here are only a very small portion of the data, but are informative because they give an indication of the accuracy of the survey responses.

Problems with derived values in Excel

The number of counties reported in the original Excel file should be consistent with the number of counties derived from the data in R.

Table 6: Some discrepancies in number of counties per species.

| Survey No. | Lic. No. | Species | No. Counties (Excel) | No. Counties (Actual) | County | No. Fish |
|------------|----------|----------------|----------------------|-----------------------|--------|----------|
| 693 | 1731 | Rainbow Trout | 1 | 2 | ANTI | 1 |
| 693 | 1731 | Rainbow Trout | 1 | 2 | PICTOU | 1 |
| 874 | 5497 | Speckled Trout | 1 | 2 | CB | 10 |
| 874 | 5497 | Speckled Trout | 1 | 2 | VIC | 7 |
| 1463 | 18267 | Speckled Trout | 2 | 1 | GUYS | 30 |
| 1464 | 18284 | Speckled Trout | 2 | 1 | HALI | 13 |
| 1467 | 18302 | Speckled Trout | 2 | 1 | HALI | 2 |
| 1468 | 18306 | Speckled Trout | 2 | 1 | HALI | 6 |
| 1469 | 18307 | Speckled Trout | 2 | 1 | HALI | 10 |
| 1671 | 23054 | Rainbow Trout | 11 | 1 | KINGS | 5 |
| 1842 | 26772 | Speckled Trout | 1 | 2 | CB | 1 |
| 1842 | 26772 | Speckled Trout | 1 | 2 | HALI | 1 |

There is quite clearly an error, for example, in the entry that states that one respondent fished in 11 counties when they actually only fished in 1.

There are also inconsistencies between whether an entry was considered for calculating effort in the original data and the value derived in R. Sometimes the difference is due to an error in the computation of the number of counties, but sometimes it appears that this variable was evaluated on a case-by-case basis. Problems with determining which entries can be used for calculating county-specific CPUE will be discussed in the recommendation section of the report.

Recommendations

Data storage

If data are stored in Excel or comma-separated value format, variable (column) names should not begin with non-letter characters or contain spaces. Best practice is to use underscores to separate words or word fragments. Beginning a variable name with a character such as “#” causes it to be replaced with “X” in R and spaces are turned into periods.

There should be restrictions on the values that certain variables can take. For example, days and numbers of fish should only be integers, and days should always have a value of 1 or greater. If a value is illegible, instead of “?”, the value should be entered as NA.

We also recommend moving to a long data format, where each catch observation becomes a unique row in the data, rather than the current wide format. The current format has lots of extra values for each combination of species and county for each respondent that are NA, which can be removed from long-format data. Here is an example of the first 8 rows of long-format data. Each row is a complete record.

```
##  survey_num lic_num res_type rep_fished rep_caught act_caught
## 1      543     31   1 day      TRUE      TRUE      TRUE
## 2      544     45     N      TRUE      TRUE      TRUE
## 3      545     52     N      TRUE      TRUE      TRUE
## 4      545     52     N      TRUE      TRUE      TRUE
## 5      546     86     N      TRUE      TRUE      TRUE
## 6      547     88     N      TRUE      TRUE      TRUE
## 7      547     88     N      TRUE      TRUE      TRUE
## 8      547     88     N      TRUE      TRUE      TRUE
##      species sp_days n_counties_act eff_calc_act county num_fish
## 1      Shad      1           1           TRUE    ANNA      4
## 2 Yellow Perch    5           1           TRUE    ANNA      5
## 3 Speckled Trout  4           1           TRUE    ANNA      6
## 4   Brown Trout  1           1           TRUE    HALI      0
## 5 Rainbow Trout  8           1           TRUE    VIC       1
## 6 Speckled Trout  5           1           TRUE    ANTI      8
## 7   Brown Trout  2           1           TRUE    ANTI      4
## 8 Rainbow Trout  1           1           TRUE    ANTI      1
```

Recording data

Basic checking for inconsistencies, such as the problems with individuals claiming to not have fished while also reporting a number of fish caught, should be applied to online survey responses.

While it is possible that the duplicate license numbers in the data analyzed in this report are erroneous, and license number should be unique for each survey response, we still recommend adding a new variable to represent survey number to link all observations from the same respondent.

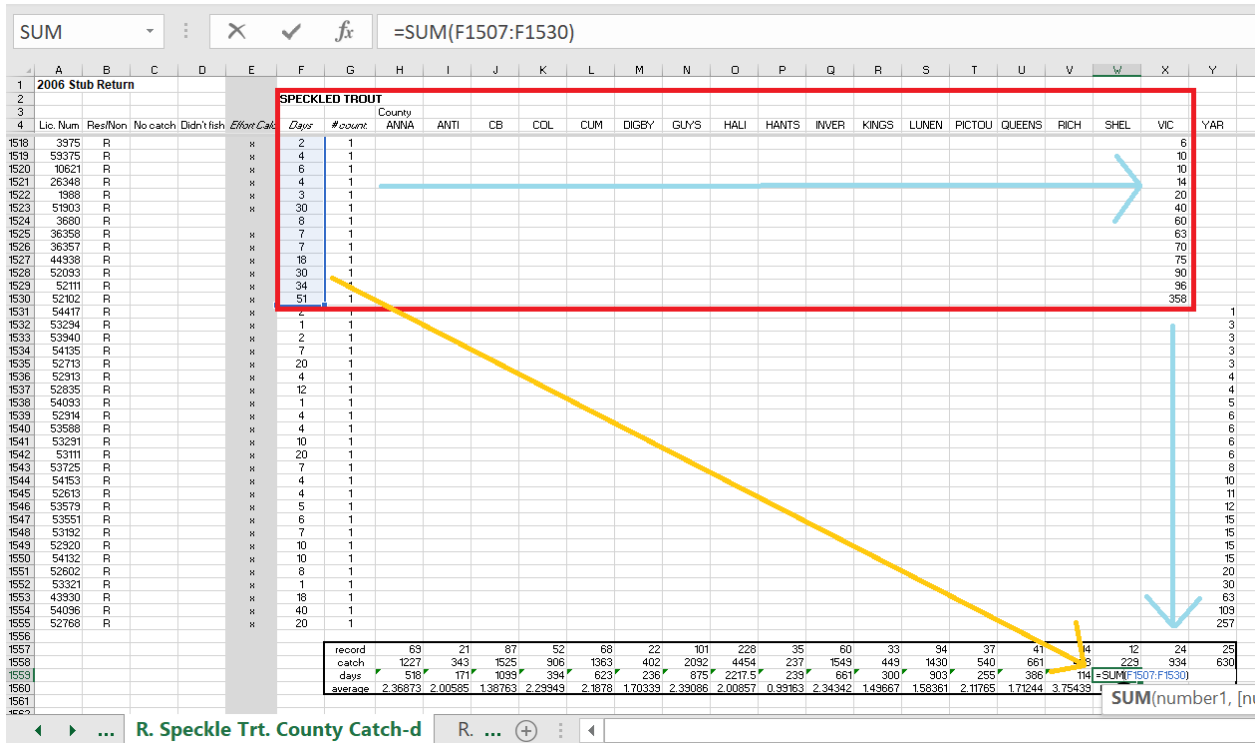


Figure 3: The cells selected for the days fished in Shelburne actually correspond to the days fished in Victoria.

Derived variables

Variables such as the number of counties fished should be computed from the data rather than entered by hand to avoid data entry errors. As shown in the previous section, calculating these values in R is not error-prone like entering values by hand is. Calculating CPUE is also more less error-prone in R than in excel. For example, the excel method caused a miscalculation of the CPUE for speckled trout in Shelburne county because the wrong section of data the data were selected for the number of days (Figure 3).

Survey design

There are three problems that come from the current survey design, where the number of fish caught is divided by species and county but the number of days fished is only divided by species. The latter two problems are explained in more detail in the following two subsections

1. Grouping different variables by different numbers of factors makes data processing somewhat more difficult, and thus potentially error-prone (particularly in the long data format, where the number of days associated with a given observation is not necessarily the number of days required to catch the number of fish in that observation). The number of counties fished for a given species and the column indicating whether the observation can be used for CPUE calculations are only required because information on fishing effort is not collected per species *per county*.
2. The data that can be used for per county CPUE calculations is only a subset of the data that can be used for per species CPUE calculations.
3. The survey design may be confusing to respondents and cause erroneous responses.

Reduced data for per county calculations

There are 3080 survey responses in the entire 2006 dataset, with 4279 catches reported (and 1043 respondents who did not report a catch). The total number of observations available for CPUE calculations (number of fish and number of days both present) is 3102. Of those, only 2164 observations (69.8%) can be used for per county CPUE calculations.

It would be beneficial to allow respondents to indicate the number of days spent fishing in each county because 14.7% of survey respondents reported fishing for at least one species in multiple counties.

Problems with interpreting survey responses

As stated previously, the decision to include or exclude each entry in CPUE calculations in the original data appears to have been evaluated on a case-by-case basis, where, sometimes, observations which should fit the requirements were excluded and observations which should not fit the requirements were included. Most of these inconsistencies do not actually correspond to whether or not the observations were actually used for CPUE calculations in subsequent sheets in the Excel file.

However, the example in Table 7 reveals how survey respondents may not fill out the survey correctly and was not included in the original CPUE calculations for that reason, but still fit the criteria for CPUE calculations in this report.

Table 7: An example of listing days per county rather than per species.

| License No. | Species | Days per Species | County | No. Fish |
|-------------|----------------|------------------|--------|----------|
| 96 | Speckled Trout | 120 | ANTI | 10 |
| 96 | Brown Trout | NA | ANTI | 4 |
| 96 | Rainbow Trout | NA | ANTI | 60 |
| 96 | White Perch | NA | ANTI | 20 |
| 96 | Yellow Perch | NA | ANTI | 15 |

The data state that the respondent spent 120 days fishing for 10 Speckled Trout in Antigonish, and an unknown amount of time fishing for other species in the same county. However, it seems very likely that the respondent actually spent 120 days in Antigonish fishing for a variety of species. Claiming that only 10 fish were caught in 120 days gives a very low CPUE for that combination of fish and county, which is likely why it was not included in the original CPUE calculations.

Only 81 respondents fished for multiple species in only one county while reporting the number of days only once (4.58% of total observations where a catch was reported), and 9 respondents fished for multiple species in only one county and reported the number of days more than once but not for each observation (0.771% of total observations where a catch was reported), so we cannot assume that all respondents intended to respond this way. The amount of data that has to be discarded because of survey respondent confusion is not large, but, as in the example in Table 7, can lead to outliers in the data.

Flexibility of survey

Only 10 species of fish are included in the current license stub surveys, but anglers may catch other species while fishing. We recommend allowing respondents to indicate other species which they catch to collect as much data as possible.

Analysis

The data can be summarized by individual survey responses, by county, by species, and by both county and species. When calculating average catch per unit effort (CPUE), there are three things to keep in mind:

- CPUE is to be calculated as the total number of fish divided by the total number of days, rather than the average of the number of fish divided by the number of days for each angler, county, or species
 - thus, the overall average should be the same when the same data is used
 - this method is used because it is less sensitive to extreme values than taking the average of many ratios
- only values of fish caught with corresponding values of days spent fishing can be used for CPUE calculations (and vice versa), meaning CPUE calculations can only ever be performed on a subset of the data, as some people reported fish numbers without reporting the number of days and a few reported the number of days but not the number of fish caught
 - because some people reported fishing but did not report a catch, we cannot conclude that failure to indicate the number of fish caught indicates that no fish were caught for the given number of days
- CPUE can only be calculated on a per-county basis for those records wherein the angler only fished for the given species in one county; thus, the by-county and by-species-and-county CPUE estimates are performed on less data than the by-species CPUE estimates

By Survey

When data is summarized by survey respondent, effort can only be calculated for individuals who reported the number of days fished for each catch. Therefore, some data is dropped from the total CPUE calculations. For example, assume an angler reported catching 10 Speckled Trout in 10 days and 10 Brown Trout in an unknown number of days. The total number of fish caught by that angler is 20, but we cannot know the total number of days spent fishing and thus cannot know the CPUE for that angler. Given the data used, the overall CPUE is 2.31 fish per days fished.

Figure 4 summarizes the number of species and counties fished by individuals and the catch, effort, and CPUE distribution across all survey respondents.

Figure 5 shows the fraction of successful anglers.

Most respondents only fished for one species and most respondents only fished in one county. Figure 6 shows the relationship between days fished and total catch and between days fished and CPUE.

By Species

The number of fish caught and number of anglers angling for each species are summarized in Figure 7.

The number of days angled for each species and CPUE are summarized in Figure 8.

The overall average CPUE for the data used in these plots is 2.3 fish per days fished.

By County

The number of fish caught and number of anglers in each county are summarized in Figure 9.

The number of days angled in each county and CPUE are summarized in Figure 10.

The overall average CPUE for the data used in these plots is 2.04 fish per days fished.

By County and Species

The number of fish caught and number of anglers for each species in each county are summarized in Figure 11.

The number of days angled in each county and CPUE are summarized in Figure 12.

The average CPUE of each species in each county are also shown in Table 8.

Table 8: CPUE per each county and species combination

| County | Brown Trout | Chain Pickerel | LL Salmon | Rainbow Trout | Shad |
|--------|-------------|----------------|-----------|---------------|-------|
| ANNA | 0.592 | 1.00 | NA | 1.450 | 1.670 |
| ANTI | 0.913 | NA | 1.000 | 0.744 | NA |
| CB | 0.439 | NA | 1.000 | 0.836 | 2.000 |
| COL | 1.320 | 3.44 | 2.590 | 1.410 | 2.130 |
| CUM | 1.320 | NA | NA | 0.875 | 0.474 |
| DIGBY | NA | 2.38 | NA | 2.270 | NA |
| GUYS | 0.654 | 5.00 | 1.000 | 1.080 | NA |
| HALI | 1.120 | 1.25 | 1.040 | 0.640 | 1.290 |
| HANTS | 0.286 | 2.26 | 0.167 | 1.120 | 2.230 |
| INVER | 0.632 | NA | NA | 0.410 | NA |
| KINGS | 0.612 | 2.89 | NA | 2.290 | 4.000 |
| LUNEN | 0.308 | 5.60 | 0.000 | 1.980 | 7.000 |
| PICTOU | 1.460 | 2.50 | NA | 1.690 | NA |
| QUEENS | NA | 1.60 | NA | 5.940 | NA |
| RICH | 1.200 | NA | NA | 4.000 | NA |
| SHEL | 0.333 | 8.31 | NA | 0.667 | NA |
| VIC | 0.800 | NA | NA | 0.470 | NA |
| YAR | NA | 2.86 | NA | 1.710 | NA |

| County | Smallmouth Bass | Speckled Trout | Striped Bass | White Perch | Yellow Perch |
|--------|-----------------|----------------|--------------|-------------|--------------|
| ANNA | 1.760 | 2.36 | 1.680 | 0.766 | 3.90 |
| ANTI | NA | 1.22 | NA | 0.357 | 3.00 |
| CB | NA | 1.58 | NA | 1.980 | 1.33 |
| COL | 0.533 | 2.36 | 0.825 | 1.650 | 2.37 |
| CUM | 2.830 | 2.19 | 0.657 | 7.330 | 2.17 |
| DIGBY | 3.050 | 1.70 | 1.000 | 0.642 | 4.79 |
| GUYS | NA | 2.41 | 0.000 | 3.600 | 3.70 |
| HALI | 2.890 | 2.02 | 1.020 | 1.310 | 1.01 |
| HANTS | 1.930 | 1.04 | 0.689 | 1.750 | 3.82 |
| INVER | 3.000 | 2.27 | NA | 3.780 | 4.29 |
| KINGS | 5.230 | 1.50 | 0.938 | 2.590 | 1.19 |
| LUNEN | 4.000 | 1.58 | 2.170 | 1.970 | 1.40 |
| PICTOU | 3.000 | 2.11 | 0.258 | 1.500 | NA |
| QUEENS | 10.600 | 1.73 | NA | 1.360 | 1.94 |
| RICH | NA | 3.94 | NA | 2.500 | NA |
| SHEL | NA | 1.67 | NA | NA | 2.18 |
| VIC | NA | 3.83 | NA | NA | 1.14 |
| YAR | 2.410 | 2.77 | 0.444 | 2.250 | 2.55 |

The overall average CPUE for the data used in these plots is 2.04 fish per days fished.

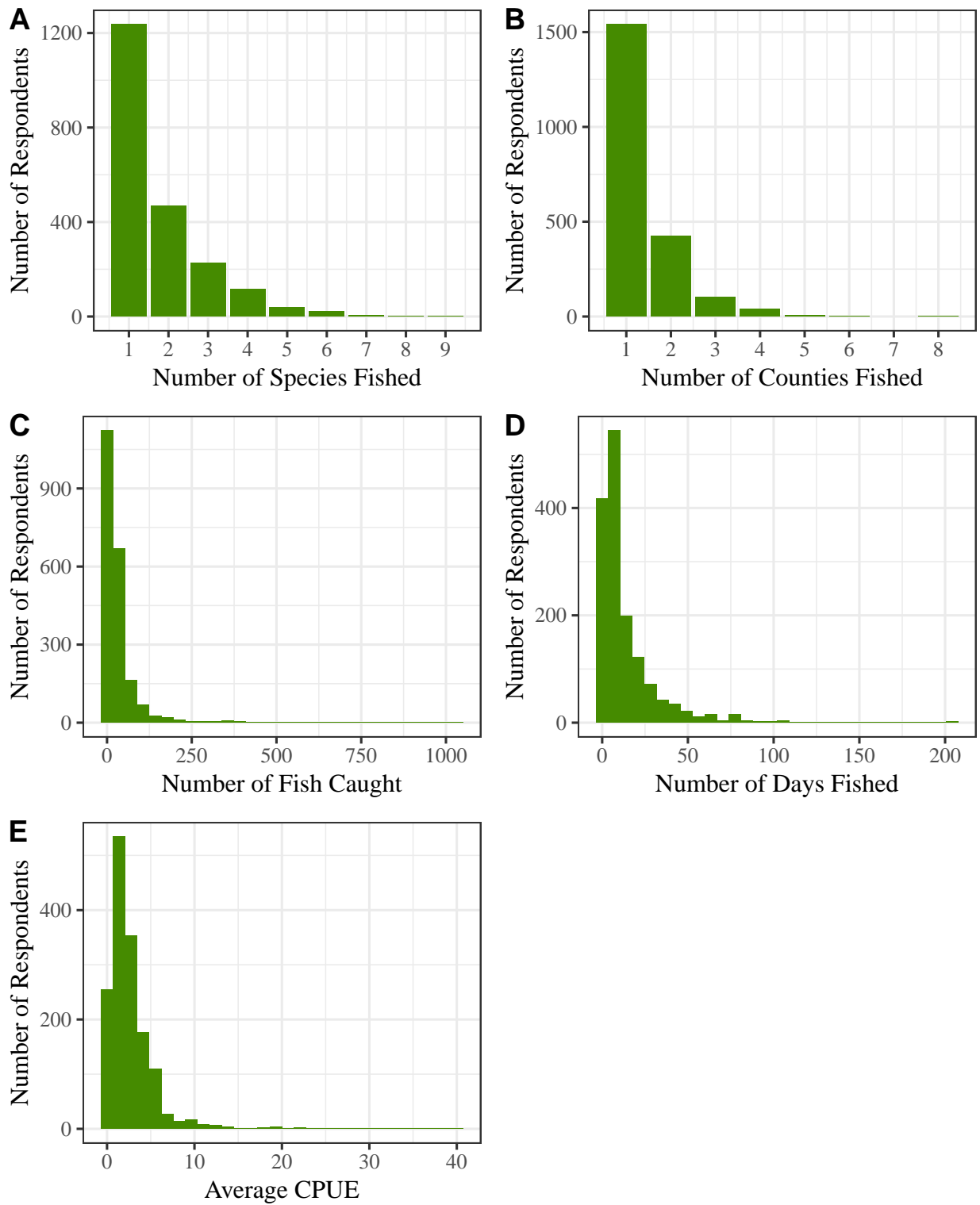


Figure 4: Histograms showing the frequencies of survey respondents for the the number of species fished for (A), the number of counties fished in (B), the number of days spent angling (C), the number of fish caught (D), and the average CPUE (E). All histograms appear to follow a Poisson distribution.

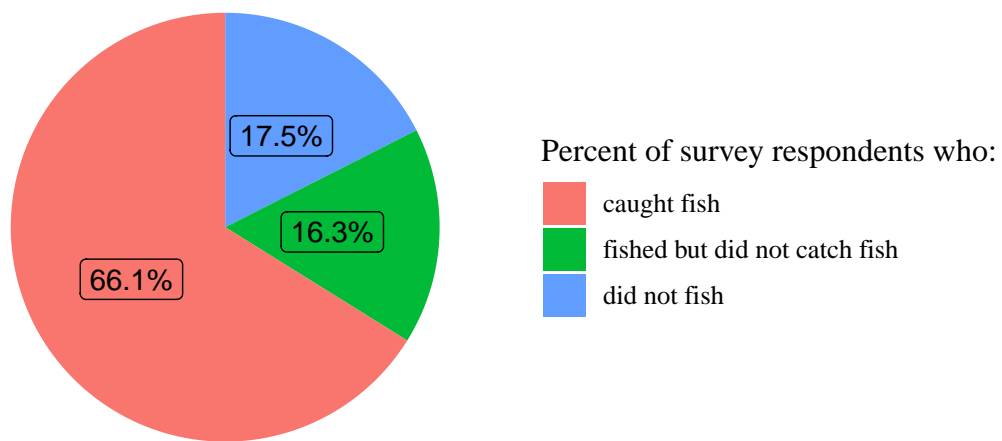


Figure 5: The proportion of survey respondents who caught fish.

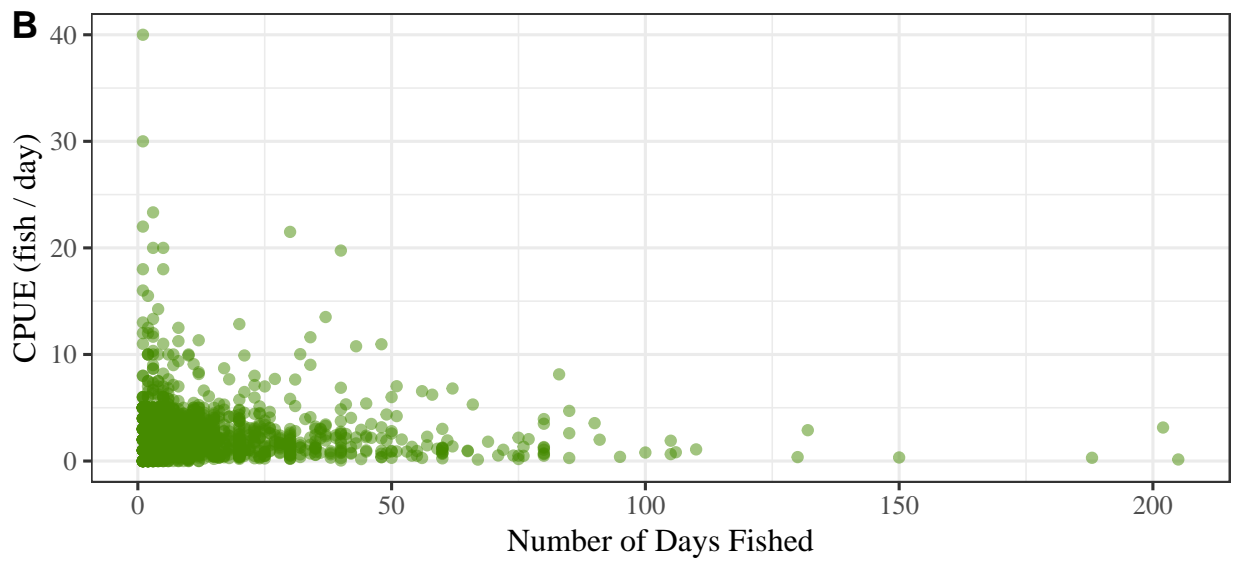
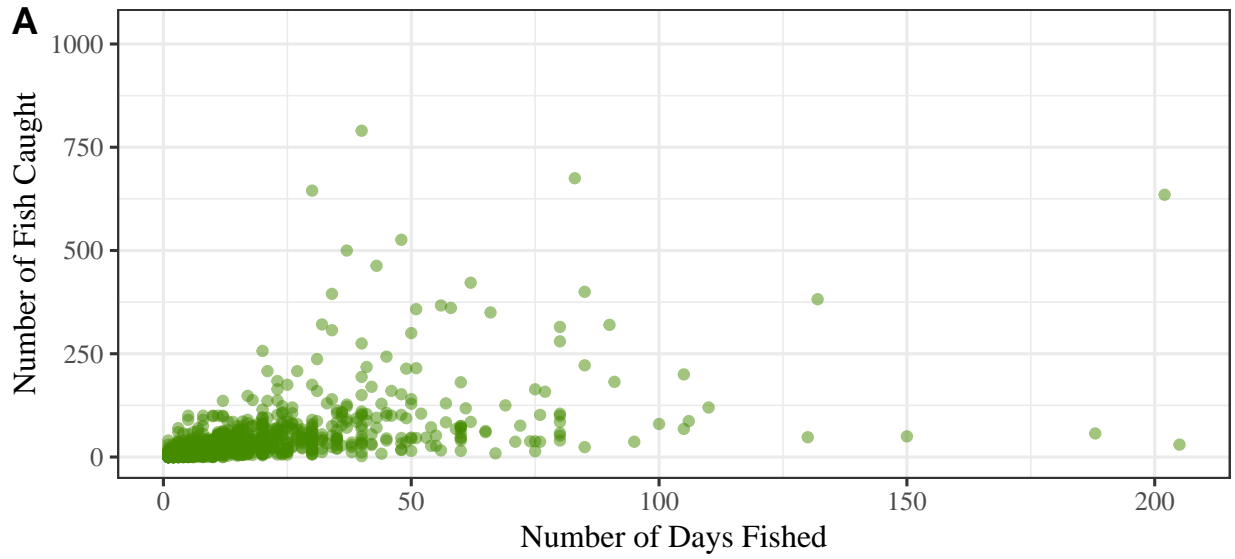


Figure 6: Scatterplots showing how many days each respondent fished and the number of fish they caught (A) and their average CPUE (B).

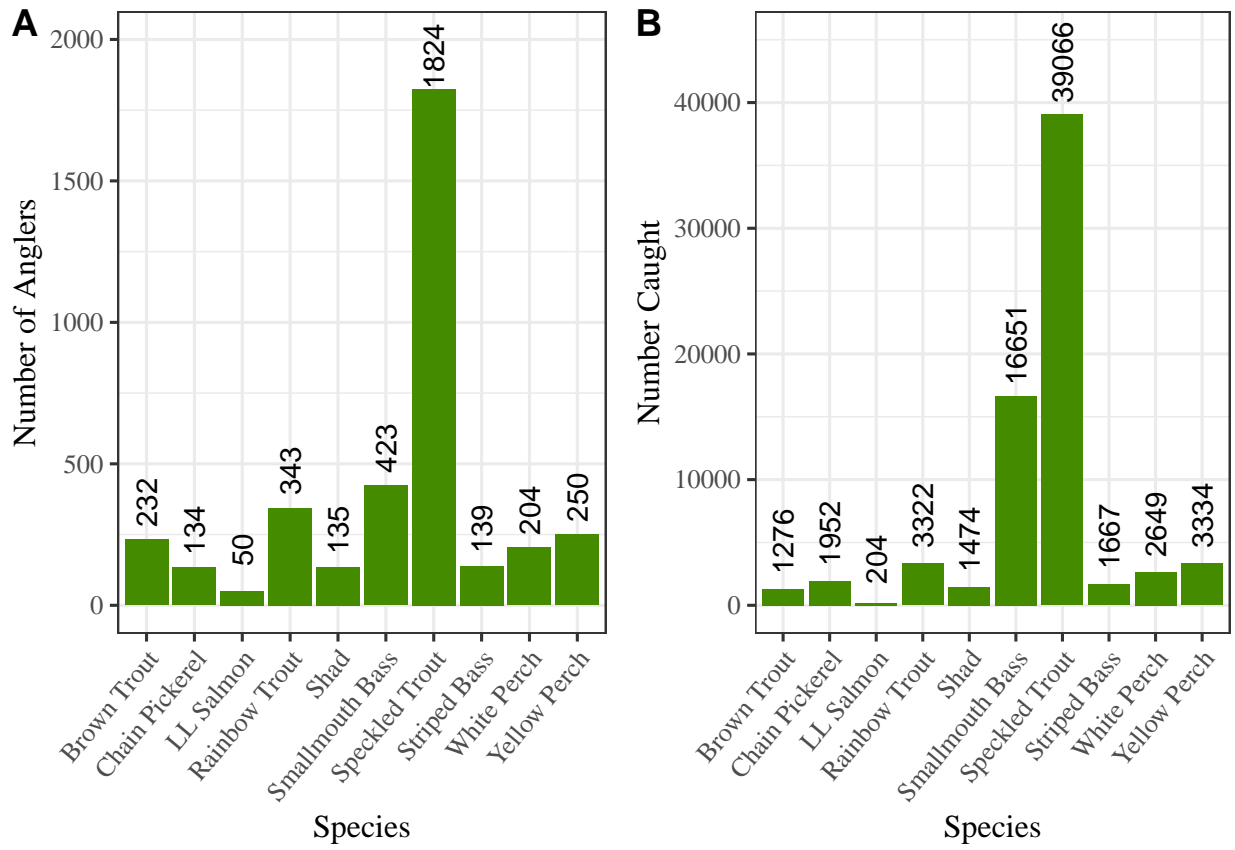


Figure 7: The number of anglers who fished for each species (A) and the number of fish of each species caught (B), calculated from the full dataset.

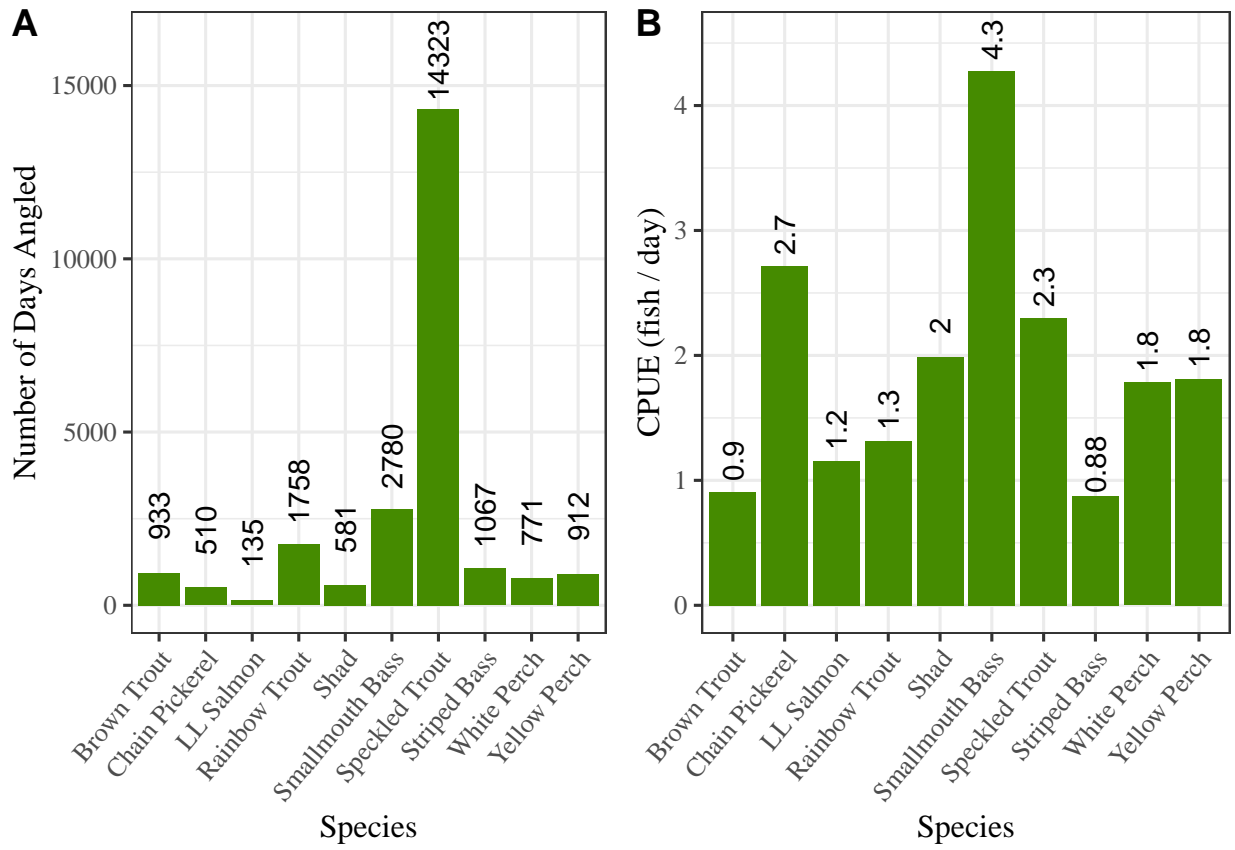


Figure 8: The number of days spent angling for each species (A) and the average CPUE for each species (B), from the subset of data which includes catch and effort information.

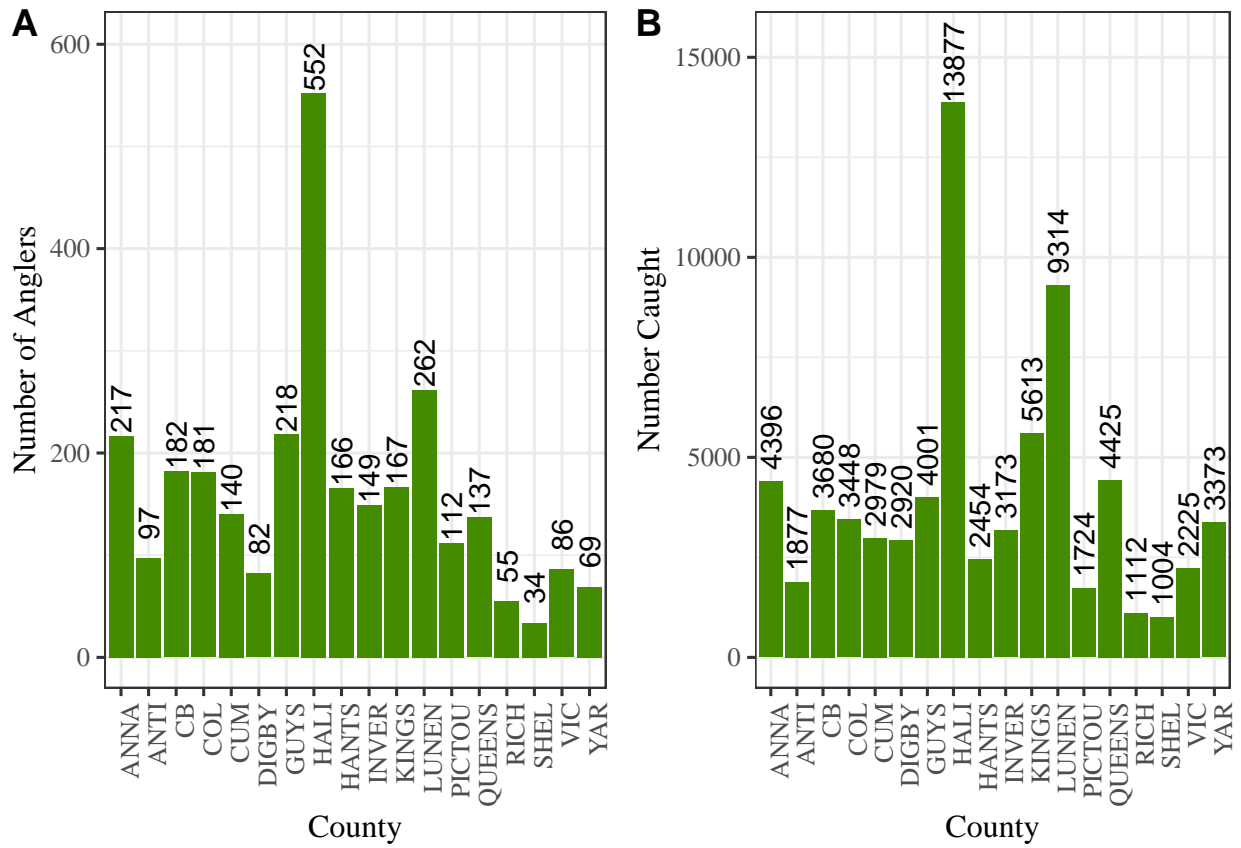


Figure 9: The number of anglers who fished in each county (A) and the number of fish caught in each county (B), calculated from the full dataset.

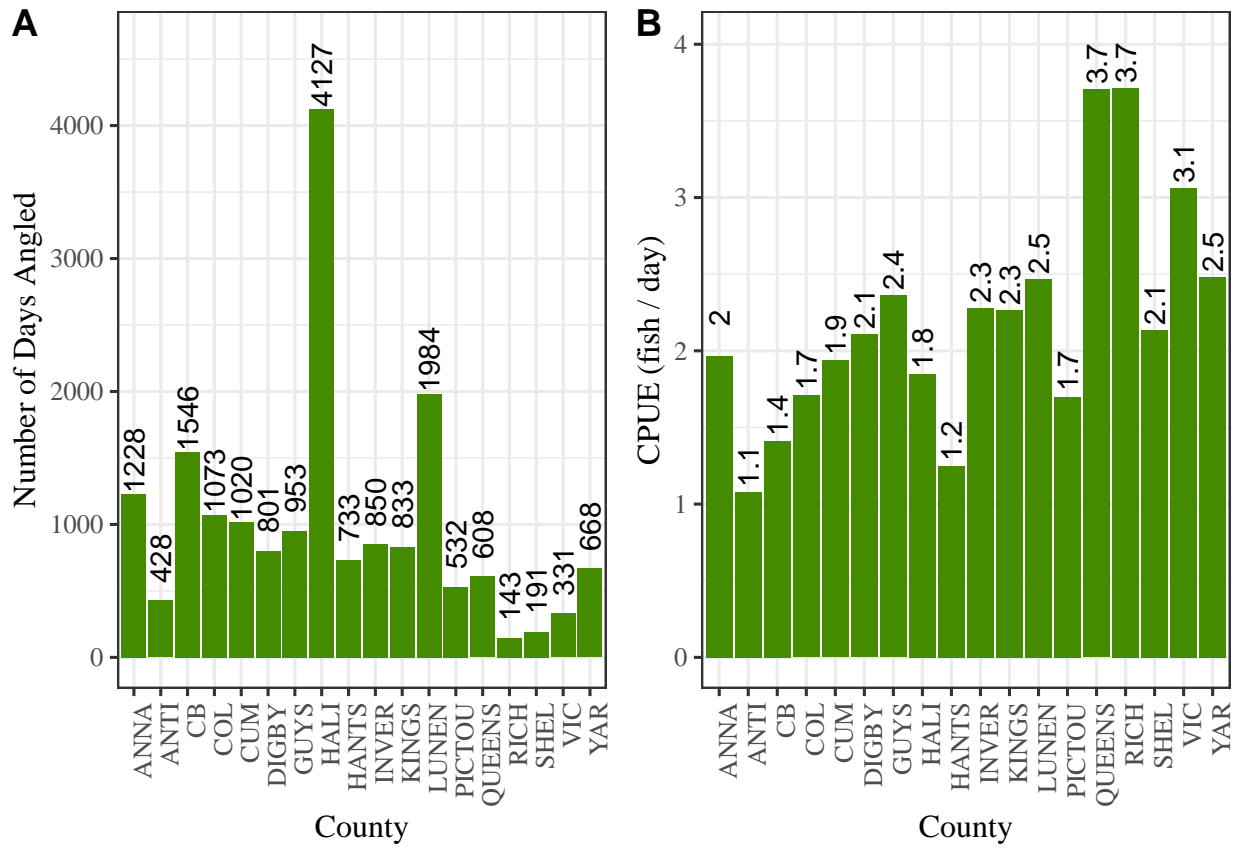


Figure 10: The number of days spent angling in each county (A) and the average CPUE for each county (B), from the subset of data which includes catch and effort information and only one county per species.

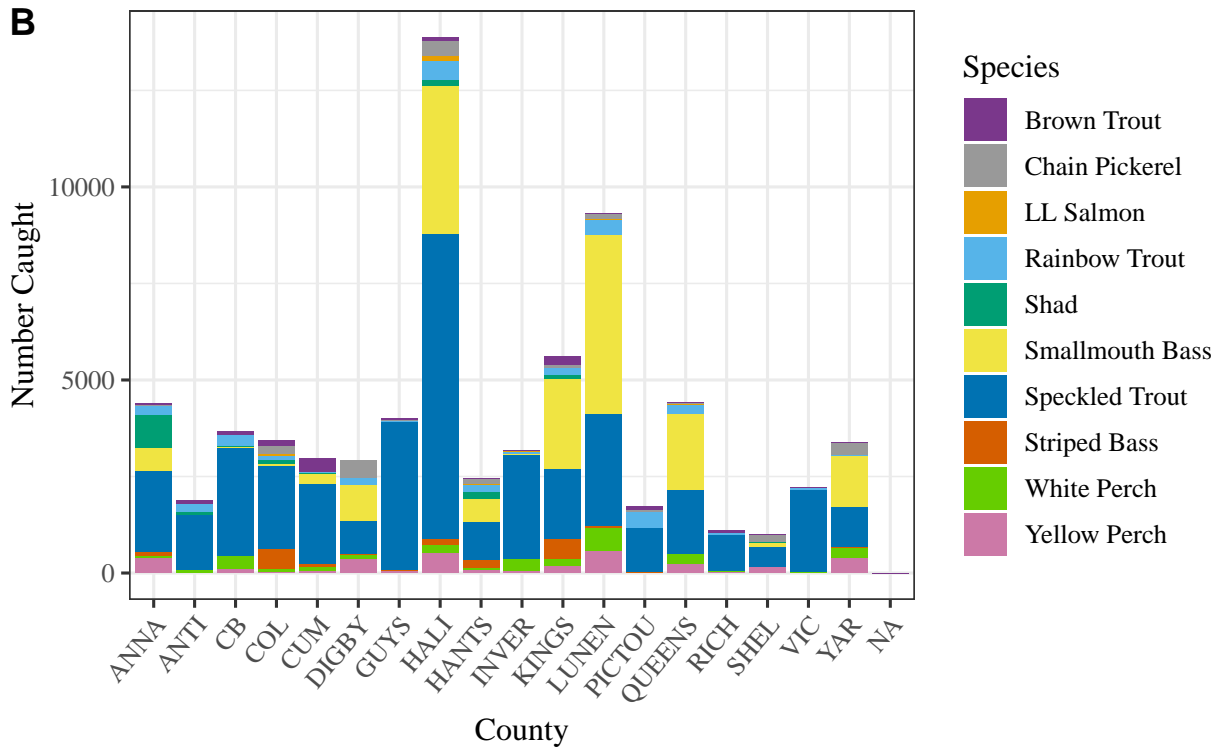
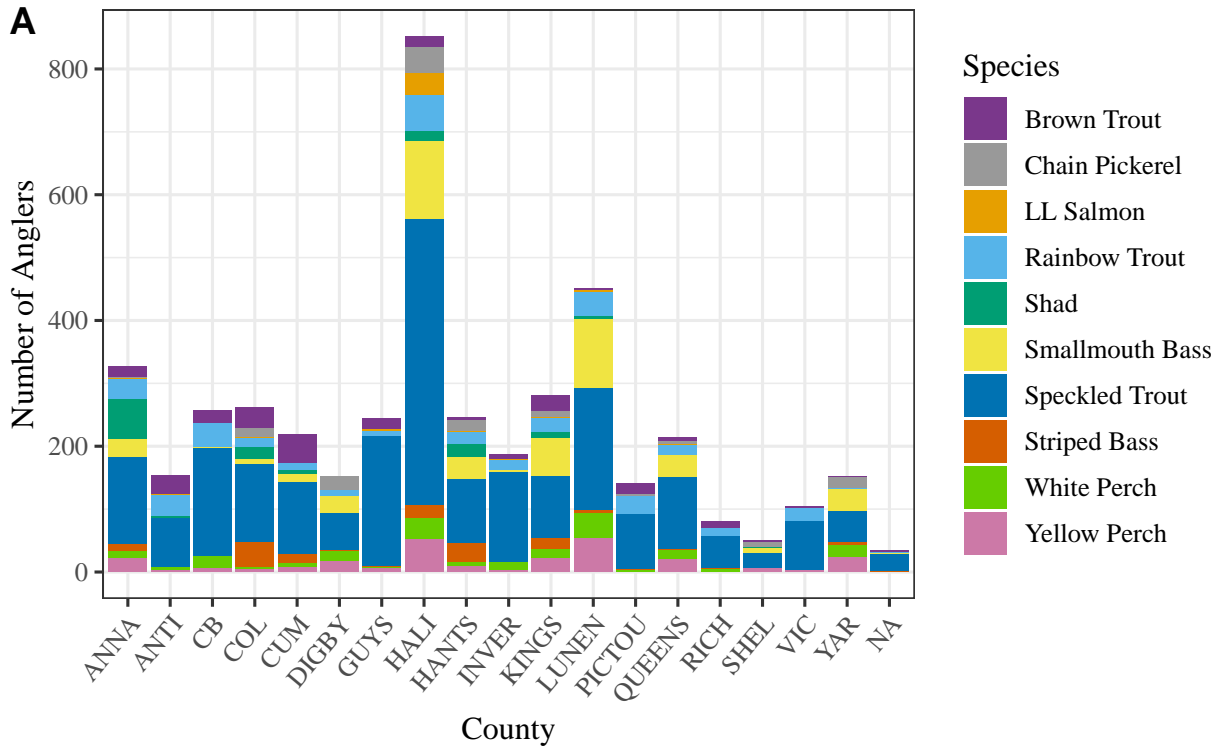


Figure 11: The number of anglers who fished for each species in each county (A) and the number of fish of each species caught in each county (B), calculated from the full dataset.

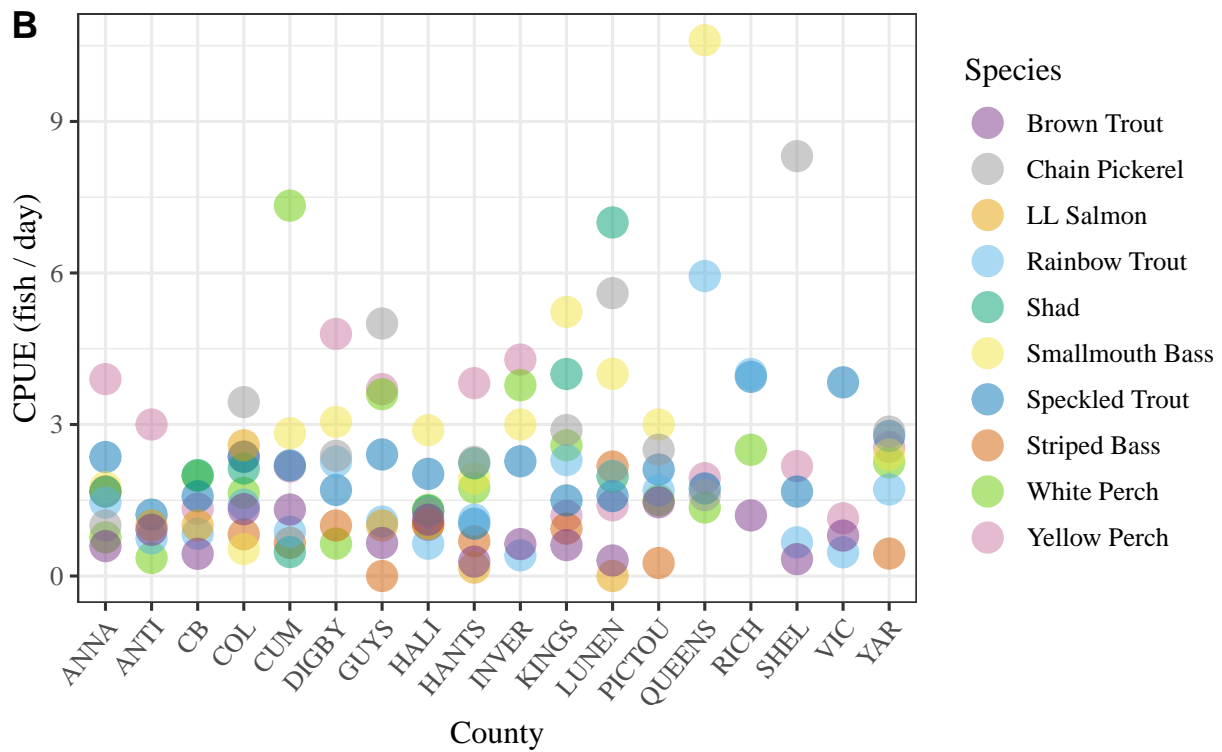
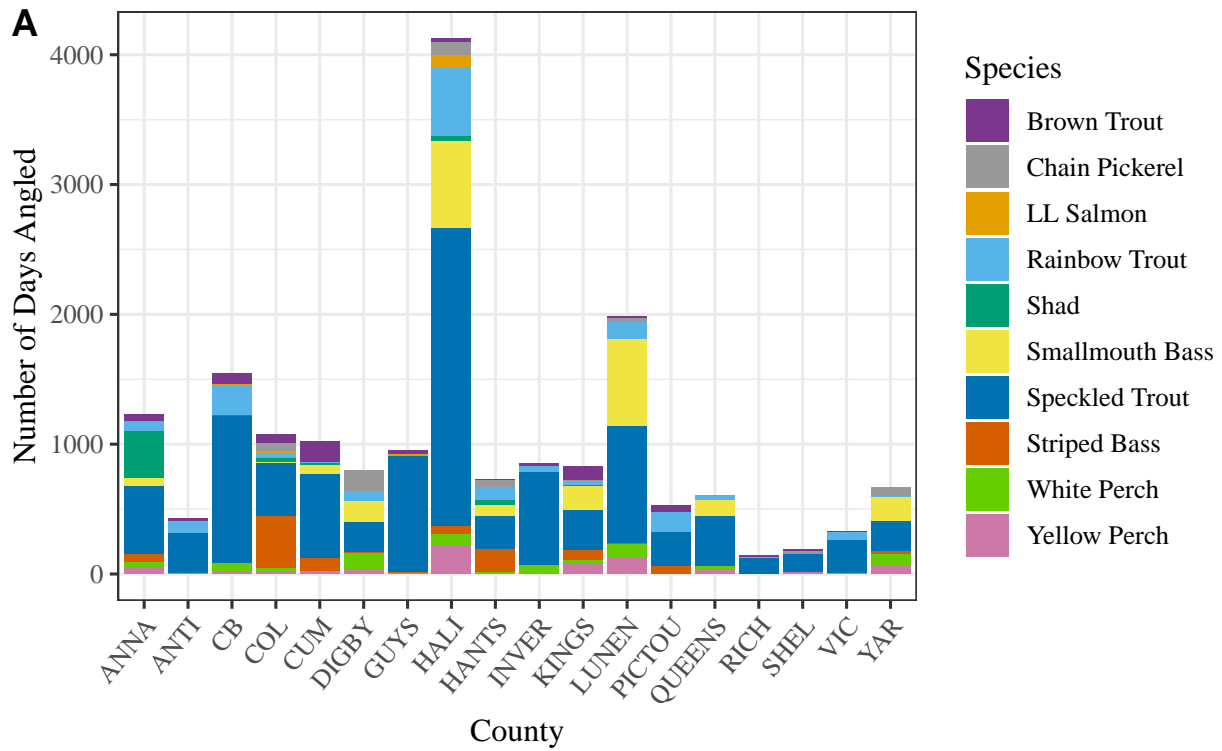


Figure 12: The number of days spent angling for each species in each county (A) and the average CPUE for each species in each county (B), from the subset of data which includes catch and effort information.