R programming, final project

This is the final project for the R-programming course in the autumn of 2016. Collaboration between students is encouraged but reports should be handed in on a per student basis. All the analysis should be reproducible and the reports are to be submitted as a R-Markdown document. The due date for the project is the December 5th and the grading is either **Pass** or **Fail**.

The data: The Icelandic groundfish survey

The data in this assignment comes from the Icelandic Groundfish Survey (IGFS). The objective of the IGFS is to improve the scientific basis of fisheries management and thus increase consistency and reliability of management advice for demersal fish stocks in Icelandic waters. The survey has been carried out annually in March since 1985, covering the continental shelf waters around Iceland with 540-600 "semi-randomly" distributed tows, more commonly refferred to as towstations (or simply stations). Biological data are collected for all fish species although sampling effort is somewhat related to economic importance of each species.

In the survey measurements of the catch composition in each tow is recorded. It is often infeasible to measure all fish that are caught thus measurements are often taken from a subsample of the total catch. Generally the following is recorded for a particular tow:

- Species: All species are identified and the total numbers/weight caught is recorded
- Length measurement: Direct measures of fish in cm. Quite often there are based on a subsample, the size of the subsample being species dependent (for some species all fish are measured, for some none).

The data above are often stored in separate tables and at the MFRI these are recorded in three separate tables:

Station table

Tee the `Station.csv` file. The table contains the following fields that store information recorded from individual tows:

- **id**: Unique station id
- **date1**: Tow start time
- **date2**: Tow end time
- **lon1** and **lon2**: Longitude of Tow start and end location
- **lat1** and **lat2**: Latitude of Tow start end end location
- **towlengt**: Tow length in nautical miles
- **depth**: Mean bottom depths in meters
- **temperature**: Bottom temperature
- **strata**: The strata of the tow

Note that each tow has a unique ID, called the station ID. This ID is recorded in all tables, which allows linking between the tables.

Subsampling table

See the `Subsampling.csv` file. The table contains information about the subsampling at each station:

- **id**: Station id
- **species**: Species code
- **n.counted**: Number of fish that were counted

- **n.measured**: Number of fish that were measured
- **n.total**: Total number of fish caught

Length table:

See the `Length.csv` file. The table contains information about fish measurements:

- **id**: Station id
- **species**: Species code
- **length**: Length class in centimeters
- **n**: Number of fish measured
- **sex**: Sex code, females (2) and males(1)
- **maturity**: Maturity code

The assignment

In this assignment catches of haddock in the groundfish survey are to be investigated. You will create:

- Maps of habitat, i.e. where can the species be found and in what quantity
- Illustrations of changes in size composition
- Calculate an index of biomass and recruitment (i.e. biomass of 1 year olds)

To produce this information at bit of data processing is required as the different parts of the survey data is contained in these three tables. First the number of fish length measured needs to be corrected due to subsampling. Secondly the biomass of each tow (i.e. catch) needs to be calculated based on a length-weight relationship. Finally the data needs to be summarised by year to calculate an index of biomass. The subsequent sections will guide you through the necessary steps of the analysis.

1. Tow locations

- From the date columns read the year of the survey.
- Investigate the location of the tow station. Choose a year, e.g. the year 2005, and plot the location of the tows from the stations table on a map of the Icelandic continental shelf. Using `geom_segment`, illustrate the start and end points of the tows.
- **Bonus**: find out how to draw arrows indicating the tow direction.

2. Subsampling table

The subsampling table records how many fish where caught and how many are length measured. When the number of fish exceed a certain number a semi-random subsample is taken from the catch and only fish in this subsample is length measured directly.

- For each station, calculate the ratio of the total number to measured fish i.e.

$$r = \frac{n.total}{n.measured}$$

  and when r is `NA` replace it with the value 1.
- Plot the average ratio by year. Note you will need to get the year from the stations table

3. Length table

The length table contains the information on the length l in cm and the number of fish of that length that were measured. The total biomass is calculated as the approximate weight of fish at length l times the number of fish in the Tow. To calculate the biomass at at length l you will need to use a length weight relationship, which is adequately described by

$$W_l = 10^{-5} l^3$$

The number measured need to be scaled using the ratio for each station you calculated in the previous step.

- Assign weight to each length in the length table
- Calculate the scaled numbers at length
- Plot the scaled length distribution by year
- **Bonus**: Plot the mean length at year.

4. Survey indices

Survey indices are used to give relative changes in biomass, and essentially serve as indicators of the effect of fisheries of the resource. A number of methods exist to calculate this index, but here we will only consider the mean catch per year:

$$I_y = \frac{\sum_s \sum_{l_{min}}^{l_{max}} B_{sly}}{N_y}$$

where I_y is the index at year y, $B_{sly} = N_{lsy} W_l$ is the biomass of length l, at station s and year y, N_{lsy} is number of fish in the same dimension, N_y the number of stations in year y and l_{min} and l_{max} represents the length range of the index.

- Calculate the total biomass index by year and illustrate
- Assuming all fish smaller that 20 cm is age 1, calculate a recruitment index by year and illustrate.
- Illustrate locations where haddock is found, e.g by plotting station location with dots relative to the amount of catch. Indicate stations with zero catch.
- **Bonus**: For fish smaller than 20 cm compare the spatial distribution of fish in years where recruitment is poor vs good

Wrap up

- Create a function that calculates the survey index, the function should accept the three tables as input.
- **Bonus**: Do a within year bootstrap on the station table to calculate the confidence limits on for the survey index. Hint: use a combination of `group_by` and `sample_n`