# Fishing Dashboard – Improving the reporting of data on fisheries

Joonas Valve, Natural Resources Institute Finland, [joonas.valve@luke.fi](mailto:joonas.valve@luke.fi)  
Maria Yli-Heikkilä, Natural Resources Institute Finland, [maria.yli-heikkila@luke.fi](mailto:maria.yli-heikkila@luke.fi)

**Abstract**

*Member states in European Union are obligated to collect, manage, and deliver data on fisheries under the Data Collection Framework. Based on this data, researcher experts offer scientific advice for the European Commission which uses the information to pursue sustainable and economically efficient fishing policy. The Finnish data colletion is managed by Natural Resources Institute Finland. This paper aims to improve the process of data validation and quality assurance as well as enhance the disseminating of the results.*

***Keywords****: Fisheries, data visualization, dashboard, R, Markdown*

## Introduction

EU fisheries management relies on data collected, managed and supplied by European Union (EU) countries under the Data Collection Framework (DCF). The framework for the collection and management of fisheries data was established in 2000 and the current form of DCF has been executed since 2008. Under this framework the Member States collect, manage and deliver a wide range of fisheries data needed for scientific advice and wider public. Natural Resources Institute Finland (LUKE) is responsible for the data collection in Finland. LUKE publish data on fisheries in their EconomyDoctor web service and Px-Web statistics database. However, the usability and the layout could be improved. This paper aims to find new possibilities to better represent the data. Another objective is to improve the data validation and quality assurance related to the data collection.

The fleet-economic performance data call requested by the European Commission contains annual information about income, expenditure, capacity, employment and landings for each commercial fishing vessel. Based on this data Scientific, Technical and Economic Committee for Fisheries (STECF) produces the Annual Economic Report (AER) on the EU fishing fleet which provides a detailed overview of the profitability of fisheries in EU region and in each Member State.



Figure 1. Workflow of data collection framework (STECF, 2019).

The Annual Economic Report on the EU fishing fleet (Carvalho, Keatinge and Gullen, 2018), published in the EU Science Hub website, has become one of the main sources of economic and social data for scientific advice on the performance of the EU fishing sector. The fishing fleet includes all vessels used in commercial marine fishing. The AER is increasingly used by scientific bodies, national administrations and international institutions. It provides a comprehensive overview of the latest information available on the structure and economic performance of fishing fleets in EU. The report covers a time period from 2008 onwards and includes information on the capacity, effort, employment, landings, income and costs of the EU fishing fleet. Projected values and nowcast estimates for the next few years are reported where possible. The economic performance of the EU fishing fleet is also reported in terms of gross value added, profits, profit margins and productivity (labour and capital). The most recent publication includes:

1. A structural and economic overview of the EU fishing fleet for the period 2008 to 2016, with trend analyses and projection estimates for the next two years
2. A regional analysis of the EU fishing fleet by major sea area: Baltic Sea, North Sea, North East Atlantic, Mediterranean & Black Sea, as well as fleets operating in Other Fishing Regions
3. A detailed structural and economic overview of each EU Member State fishing fleet, including qualitative economic performance assessments. (Carvalho et al., 2018.)

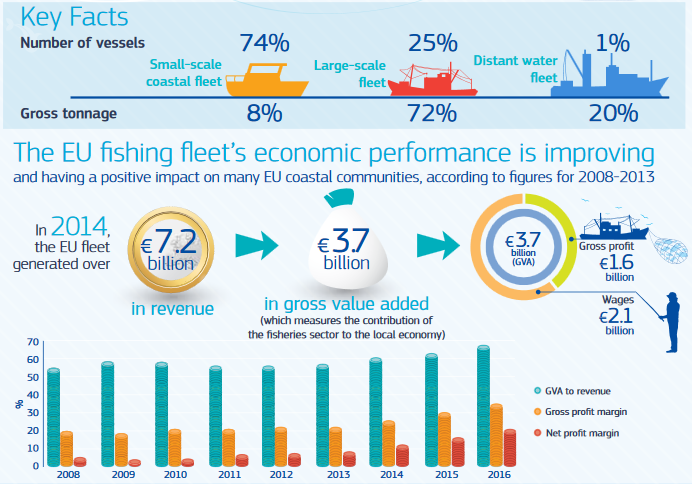


Figure 2. Key facts of the EU fishing fleet (Carvalho et al., 2018)

A project called ‘the Fishing Dashboard’ is an attempt to better visualize and communicate the Finnish data on fisheries. It provides an overview of the Finnish data collected for the fleet-economic performance data call. The initial purpose was to develop an analytical and visual tool for the statistical production chain, especially for data validation and quality assurance. However, the Fishing Dashboard can also serve as a platform for disseminating results in a well-formatted way. The dashboard is in early stages but gives a raw illustration on how to represent data using the *flexdashboard* (Iannone, Allaire and Borges, 2018) package. The Fishing Dashboard includes visualization of the annual data as well as searchable data tables. It also contains simple spatial analysis of the value of landings of finnish fishing fleet.

R Markdown is a file format for making dynamic documents with R. It was first introduced in the *knitr* package (Xie, 2017) in early 2012 which introduced the idea to embed code chunks (R or other languages) in Markdown documents. R Markdown enhance the usability of the original Markdown language by combining it with Pandoc. In short, knitr executes the embed code and converts R Markdown to Markdown and Pandoc renders the output to chosen format (Xie, Allaire and Grolemund, 2018). R Markdown enables large number of final outputs including documents and reports in different formats (such as pdf, word or html), notebooks with integrated code chunks, presentations (HTML5 and PowerPoint), dashboards, journal articles, or even books and websites (Xie al., 2018). The *rmarkdown* package (J. Allaire, Xie, McPherson, et al. 2019) was first created in 2014 and has since steadily evolved into a relatively complete ecosystem for authoring documents.

In their article Bion, Chang and Goodman (2017) point that R is one essential tool for doing data science at Airbnb. It has a role in all of the main focus areas, including data wrangling, predictive modeling and statistical testing, visualisation and reporting. R is widely used in all stages of the workflow from exploratory data analysis and modeling to reporting and sharing results with collegues and business partners. Data scientists working at Airbnb have a wide variety of duties including building data pipelines and internal data tools as well as creating reports and dashboards for sharing insights and results. Bion et al. (2017) mention several packages to enhance data wrangling (*tidyr* and *dplyr*) and visualisation (*ggplot2*) in addition to reproducible workflow (*rmarkdown*). In addition to static graphics, packages such as *plotly* and *leaflet* are often used in creating interactive visualizations or dashboards. Airbnb uses R Markdown to document all analyses with using R. It makes it easy to share code and visualizations in a single document with bussiness partners or the public. Reports also go trough a peer review process where methodologies and code style are assured before publishing and shared with business partners. The final reports and source code are shared in an internal website. This workflow allows data scientists at Airbnb to learn context and analytical techniques from collegues and previous projects. It also simplifies the coordination within the company as everyone has access to the latest research and documentation. The workflow is entirely reproducable and employees are able to reuse code from previous work. Authors conclude that none of this would be possible without the common understanding of R Markdown across the team. (Bion et al., 2017.)

Another example of the benefits of using a portable workflow is to develop an internal R package. It is a convenient way to make sure that employees are using the same tools for similar tasks and methodlogies are consistent across projects. It is also an efficient way to make coherent look for reports and visualisations within organisation for example. Airbnb has branded themes and scales to standardise their data visualisations that can be found on Github (Bion et al., 2017). BBC has also developed a R package called *bbplot* which provides helpful functions for creating and exporting graphics made in ggplot in the style used by the BBC News data team (BBC, 2019).

This paper was written entirely with R Markdown using R Studio, a free and open-source integrated development environment for R. The word document provided by the Nordic Statistical Meeting (NSM) organisers is used as a template to follow the formatting guidelines. This can save a lot of time when formatting text and figures to match the guidelines and allows more focus on the content. R Markdown also enables option to embed executable code chunks within the document. This makes completely reproducable workflow possible because the output is always produced from the source code. Overall it brings closer the analysis and reporting of the results. R Markdown also enables option to use BibTex which is a reference management software typically used with LaTeX.

The related presentation is also produced by making use of R Markdown and particularly the xaringan package (Xie, 2019). It allows an easy way to represent interactive objects with HTML5 presentation format. This means that it is possible to demonstrate dynamic maps with htmlwidgets for example. The second chapter will take a closer look on the software and tools used in this project. Chapter 3 suggest a few general ideas how to make use of R in reporting. It will also go trough an example where finnish landings data is linked with georaphic information enabling geospatial analysis of the data. The last chapter will summarise and discuss the main observations troughout the project.

## Software and Tools

R is one of the most popular software for data manipulation, statistical computing and graphics. R provides a wide variety of statistical and graphical techniques, and is highly extensible. R can be regarded as an implementation of the S language which was developed at Bell Laboratories by Rick Becker, John Chambers and Allan Wilks. For computationally-intensive tasks R can exploit more general-purpose programming languages such as C and C++. R is available as Free Software under the terms of the Free Software Foundation’s GNU General Public License in source code form. (Venables et al., 2009.)

Venables et al. (2009) desbribe R environment as an integrated suite of software facilities for data manipulation, statistical computing and graphics which includes:

* an effective data handling and storage facility,
* a suite of operators for calculations on arrays, in particular matrices,
* a large, coherent, integrated collection of intermediate tools for data analysis,
* graphical facilities for data analysis and display, and
* simple and effective programming language which includes conditionals, loops, user-defined recursive functions and input and output facilities.

The document format R Markdown has become the general document format among r users. It was first introduced in the *knitr* package (Xie, 2017) in early 2012 which suggested the idea to embed code chunks (R or other languages) in Markdown documents. Markdown is a lightweight markup language with plain text formatting syntax that was created by John Gruber and Aaron Swartz in 2004. John MacFarlane has created a package named Pandoc (<http://pandoc.org>) which converts Markdown documents to a large variety of output formats. It also substantially enriches the syntax of Markdown language and enables more types of elements. For example the original markdown language did not have syntax for tables, footnotes, math expressions or citations. Practically R Markdown stands on the shoulders of knitr and Pandoc. Knitr executes the embeded computer code converting the R Markdown format to original Markdown and pandoc renders the Markdown to the chosen output format. (Xie et al., 2018.)

The rmarkdown package (Allaire et al., 2019) was first created in 2014 and has since steadily evolved into a relatively complete ecosystem for authoring documents. R Markdown enables large number of final outputs including documents and reports in different formats (such as pdf, word or html), notebooks with integrated code chunks, presentations (HTML5 and PowerPoint), dashboards, journal articles, or even books and websites (Xie al., 2018).

R comes with a sufficient amount of built-in functions and packages for the most common and basic tasks but the real power lies in the packages developed by the community. Tidyverse (Wickham and Grolemund, 2017) is a textbook example of what a coherent system of packages can achieve. It has become the most common tool for data science among R users and newcomers. The tidyverse is a system of packages developed for data manipulation, exploration and visualization that share a common design philosophy (Wickham and Grolemund, 2017). The concept of ‘tidyness’ was introduced by Hadley Wickham (2014) with a goal to make data cleaning as easy and effective as possible. Tidy datasets has a consistent structure; each variable is a column and each observation is a row which makes them easy to manipulate, model and visualize (Wickham, 2014). Wickham and Grolemund (2017) state that tidyverse intends to make working with data more efficient and productive by enabling a consistent and reproducable workflow that is easy to share with collegues and to communicate the results with the audience.

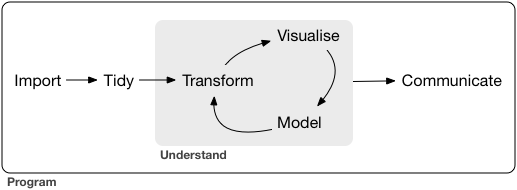


Figure 3. Workflow of a typical data science project (Wickham and Grolemund, 2017).

Dashboards are particularly common in business-style reports as they can be used to highlight brief and key summaries of a report. The layout of a dashboard is often grid-based, with components arranged in boxes of various sizes and formats. The fishing dashboard uses *flexdashboard* (Iannone et al., 2018) package to merge objects together. R Markdown language supports dozens of static and dynamic output formats which can be used with the dashboard. One advantage of flexdashboard is that all components are intelligently re-sized and adapted for the display in use (for mobiles and tablets for instance). It is also possible to specify the row or column-based layout to fill your needs. One example is the storyboard which offers explicit layout to present sequence of visualizations and adds more space for complementary commentary in form of text annotations. Flexdashboard can function with a wide variety of components including HTML widgets, R graphics, tabular data and value boxes. It also offers an option to integrate Shiny[[1]](#footnote-1) which provides endless possibilities to make elegant interactive visualizations and web applications with R. (Xie et al., 2018.)

## The Fishing Dashboard

Economic data collection is based on hierarchical multi-stage survey that combines information from different data sources. Main sources are the central control register on commercial fishery (includes fishery catch data, fishing vessel register, first hand sales of quota species), structural business and financial statement statistics from Statistic Finland (SF) and additional account surveys for coastal fisheries and trawlers conducted by Natural Resources Institute Finland.

The examples presented in this chapter utilise many different packages and key packages are reviewed here briefly. The *ggplot2* package (Wickham et al., 2019) is part of the tidyverse ecosystem and it provides a system for making elegant graphics and data visualizations using the grammar of graphics introtruced by Wilkinson, Anand and Grossman (2005). *Ggplot2* extends the idea of the grammar by building graphs from multiple layers of data (Wickham, 2010). The *DT* package (Xie, Cheng and Tan, 2019) provides an R interface to the JavaScript library DataTables. With *DT* package R data objects (matrices or data frames) can be displayed as tables on HTML pages, and DataTables provides filtering, pagination, sorting, and many other features in the tables.

The *mapview* package (Appelhans et al., 2019) provides functions to quickly and conveniently create interactive visualisations of spatial data. It is based on the *leaflet* package (Cheng, Karambelkar and Xie, 2018) which is a common open-source JavaScript library for interactive maps. *Mapview* was developed to fill the gap of quick dynamic mapping to examine and visually investigate both aspects of spatial data, the geometries and their attributes. *Mapview* also imports *sf* package (Pebesma, 2019) which offers support for simple features, a standardized way to encode spatial vector data. *Sf* package is used to link landings data with geometrical information in the spatial example.

The main objective of this project was to revise the stage of final data validation and quality assurance with simple data visualisations. The entire process of data collection involves multiple different stages, including stratified estimation method, and can be quite complex. These figures do not address the chosen methdologies but rather just investigates the results. Their purpose is to determine that there are no systematic errors in the production chain and the latest annual data values are feasible for reporting. This section propose few simple examples how to represent data with R. Figures 4 and 5 presented in this chaptes are generated using R with R Studio and the ggplot2 package.

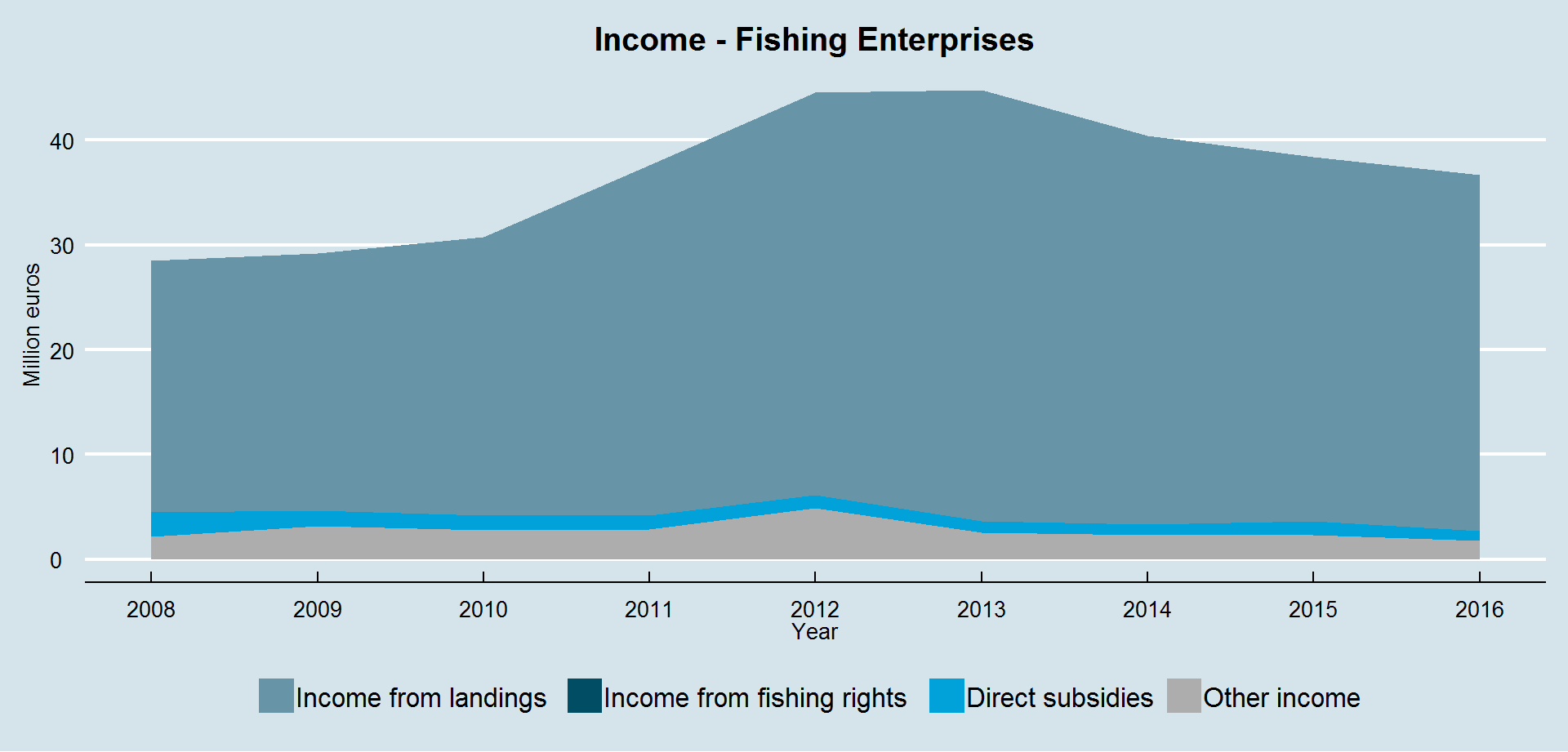


Figure 4. Income of fishing enterprises in Finland

Figure 4 shows the income of Finnish fishing enterprises during the years 2008 - 2016. Not surprisingly, practically all income comes from fish sales (landings). Landings refers to the part of fish catches that are sold. These types of annual time series graphs can be useful to determine that there are not inconsistencies in the final data before reporting the results

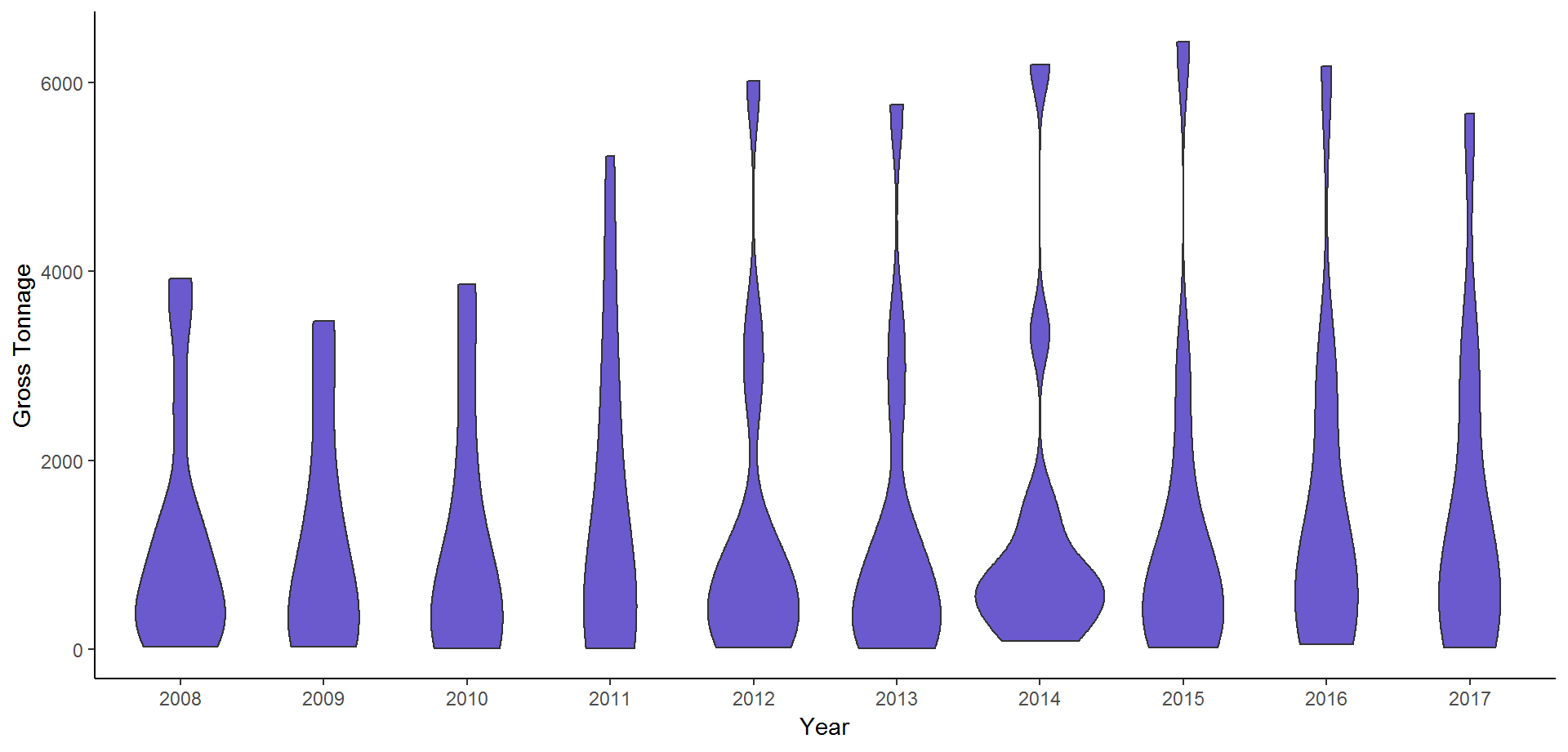


Figure 5. Capacity of Finnish fishing fleet in gross tonnage (GT)

Figure 5 express the capacity and structure of the Finnish fishing fleet from 2008 to 2017 with violin plot. The capacity is measured by the total mass of the fleet (gross tonnage). Another way to observe capacity is to look at the power of the vessels in kilowatts. Violin plots are similar to box plots but also show the probability density of the data at different values. It is often useful to see the full distribution of the data. Figure 5 gives a quick yet effective overview of the structure of the Finnish fishing fleet during the 10 year time period.

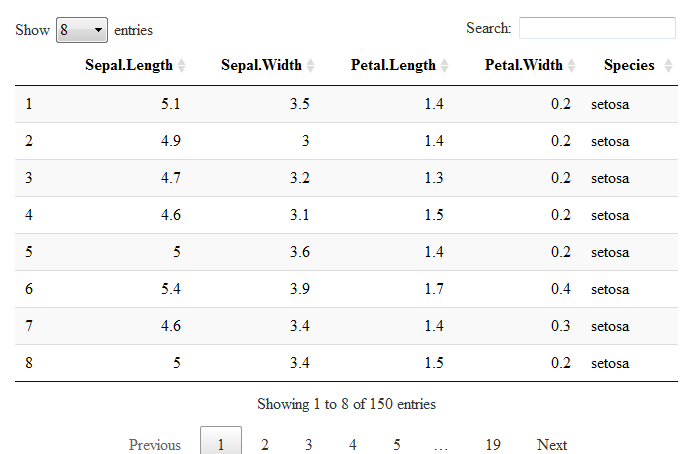


Figure 6. Example of interactive data table

Figure 6 shows a simple example with the built-in *iris* data set using the *DT* package. This figure was developed by one line of code with no customazation to the default settings. It is possible to modify the apperance of the table by using CSS and bootstrap. By default the datatable function produces search option to quickly browse trough the data but one could also add filters to colums. They are automatically adjusted to numerical and categorical variables so user is able to filter rows according to a certain range of values or classes such as species in this example. The *DT* package provides an option to represent tabular data in a well-formatted way with minimal effort. Interactivity and pre-defined filters can also be added with ease.

### Spatial analysis of Landings data

This example will consider the annual Finnish landings data from 2017. The focus is on larger scale vessels, trawlers, that mainly targets pelagic fish such as baltic herring and sprat. The fish prices by species are used to specify also the economic value of landings. Finally the data is linked with statistical regtangles enabling geospatial analysis of the data. There are few R packages that are essential in this example: *sf* for reading in and dealing with spatial data and *mapview* to conviently create interactive visualisations of spatial data with or without background maps.

Fisheries are obligated to report their catches and landings with log-book or coastal landing declaration. Log-books are used by larger vessels, mainly trawlers, whereas small-scale vessels (fishing vessels up to 10 meter) use mainly coastal landing declarations. Catches are reported by species and log-books also contain geographic information. This example concentrates on to the log-book data as it enables an effortless link to the statistical regtangles (StatRecs) used by International Council for the Exploration of the Seas (ICES).

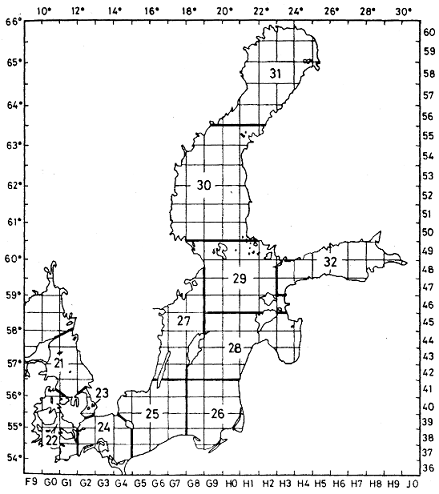


Figure 7. ICES statistical rectangles of the Baltic Sea.

The value of landings is estimated by bringing together landings and fish prices. The first-hand commercial buyers of species regulated through fishing quotas (salmon, Baltic herring, sprat and cod) have been obliged to make purchase notifications for each batch of fish purchased. The price information for these species is calculated from the purchase notifications made. The average prices of fish were calculated as averages weighted with volume purchased. The price information for species other than those covered by quotas is calculated from samples of purchasing information given by bigger fish wholesalers. The average prices of these fish are calculated as averages weighted with volume from purchasing information given by fish wholesalers. The data collected from wholesalers is highly comprehensive.

ICES standardise the division of sea areas for statistical and biological analysis. Each rectangle is approximately 30 nautical miles by 30 nautical miles. Rectangles are used for the gridding of data to make simplified analysis and visualization. The grid covers the area between 36°N and 85°30’N and 44°W and 68°30’E. ICES statistical rectangles have been in use since the 1970’s, and were first officially referenced in CM document 1977/Gen:3 ‘ICES Statistical Rectangle Coding System’ (Figure 7). Many different reference layers, including statistical rectangles for mapping spatial data can be downloaded as digital shape files from the Spatial Facility in ICES website (<http://gis.ices.dk/sf/>). After the reference layer is obtained the landings data can be attached to produce plots and maps for spatial analysis.

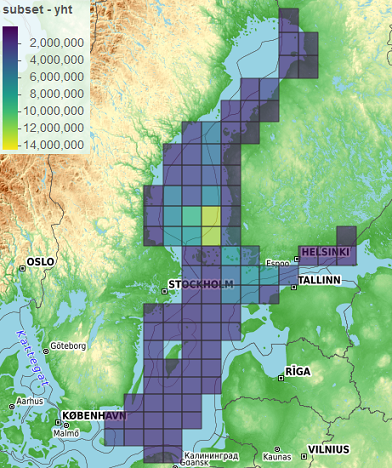


Figure 8. Total Catches (kg) of Finnish large-scale vessels in 2017 by ICES rectangles.

Figure 8 shows the total cathces in kilograms from 2017 declared trough log-books. The scale of the catches varies from 0 (dark blue) to nearly 15 million (deep yeallow) kilograms. This data does not include cathces of the small-scale vessels that normally operate closer to seashore. The values shown in this figure includes all fish species summed together. However, it is easy to examine the structure of the landings in more detail within rectangles, like shown in figure 9. It is also possible to completely change the look and feel of the map by choosing different background map.

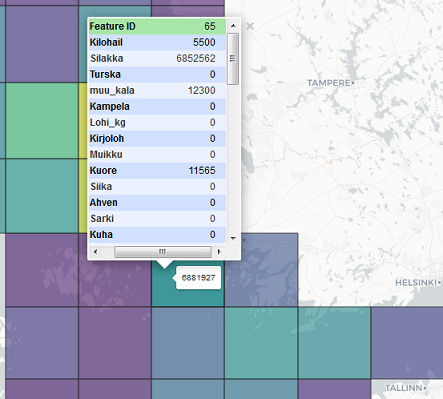


Figure 9. Mapping landings data by species to ICES rectangles.

## Discussion

EU legislation obliges Finland to gather detailed data on the Finnish fisheries industry. The EU programme for fisheries data collection seeks to gather biological, economic, and statistical data on fisheries to support the Common Fisheries Policy. The European Commission uses the data obtained through the fisheries data collection programme for the planning of fishing policy and the monitoring of it’s impacts. Our project ‘the fishing dashboard’ aims to advance the visualization, reporting and data validation of the economic fishing data collected by LUKE.

R is a free software environment aimed for data manipulation, statistical computing and graphics. R is usually used with a free and open-source integrated development environment called R Studio. R Markdown is a file format for making and authoring dynamic documents with R. It provides a fully reproducible and portable workflow. This means that the output is always compiled from the source code and it supports multiple types of output formats.

In this paper we introduced some preliminary experiments of how to represent and visualize data with R. Examples were mostly produced by using packages *ggplot2*, *DT* and *mapview*. More effort is needed to define suitable visualizations for the necessity of data collection framework and for general reporting of results. We have discussed on the possibility of producing internal r package which would provide functions for the most common visualizations as well as consistent style and layout for graphs. The *DT* package provides an option to represent dynamic tabular data in a well-formatted way with minimal effort. With *mapview* package dealing with spatial data and creating interactive visualisations of spatial data with maps is easy. *ggplot2* provides an option to revise the currently in use annual time series graphs in data validation. If needed, interactivity can be added to graphs using the *plotly* package.

We will continue our work with the fishing dashboard and explore the possibilities of communicating the data. Future work will consider if it is sensible to link data sources in a more general and consistent way to the dashboard. These could include one or both of the Natural Resources Institute Finland’s data platforms called EconomyDoctor web service and Px-Web statistics database. R has a package called *pxweb* that can interact with all PX-Web/PC-Axis APIs to fetch information about the data hierarchy, extract metadata and extract and parse statistics to R data.frame format. This should make importing data from Px-Web straight forward.

We should also consider if the dashboard or parts of it can be added to the current EconomyDoctor web service for better visual representation of the data. We have made some early testing on it that looks fairly promising. The EconomyDoctor is heavily based on the SAS programming language so there are limitations on how it can function with parallel to R.

## References

Allaire, J., Xie, Y., McPherson, J., Luraschi, J., Ushey, K., Atkins, A., Wickham, H., Cheng, J., Chang, W. and Iannone, R. (2019). *rmarkdown: Dynamic Documents for R*. <https://rmarkdown.rstudio.com>.

Appelhans, T., Detsch, F., Reudenbach, C. and Woellauer, S. (2019). *mapview: Interactive Viewing of Spatial Data in R*. <https://CRAN.R-project.org/package=mapview>.

BBC (2019). *bbplot*. GitHub repository, <https://github.com/bbc/bbplot>.

Bion, R., Chang, R. and Goodman, J. (2017). How R Helps Airbnb Make the Most of its Data. *The American Statistician*, 72(1), pp.46-52.

Carvalho, N., Keatinge, M. and Gullen, J. (2018). *The 2018 Annual Economic Report on the EU Fishing Fleet*. [ebook] Publications Office of the European Union. Available at: <http://publications.jrc.ec.europa.eu/repository/bitstream/JRC112940/kj-ax-18-007-en-n.pdf> [Accessed 25 Jun. 2019].

Cheng, J., Karambelkar, B. and Xie Y. (2018). *leaflet: Create Interactive Web Maps with the JavaScript ‘Leaflet’ Library*. <https://CRAN.R-project.org/package=leaflet>.

ICES (1977). *ICES Statistical rectangle coding system*. <https://www.ices.dk/sites/pub/CM%20Doccuments/1977/Gen/ICES%20Statistical%20Rectangle%20System%20Gen0377.pdf>.

Iannone, R., Allaire, J., and Borges, B. (2018). *flexdashboard: R Markdown Format for Flexible Dashboards*. <https://CRAN.R-project.org/package=flexdashboard>.

Pebesma, E. (2018). *sf: Simple Features for R: Standardized Support for Spatial Vector Data*. <https://CRAN.R-project.org/package=sf>.

Scientific, Technical and Economic Committee for Fisheries (STECF) (2019). *STECF - European Commission*. <https://stecf.jrc.ec.europa.eu/>. [Accessed 28 Jun. 2019].

Venables, W., Smith, D. and R Development Core Team (2009). *An introduction to R*. Network Theory Limited. <https://cran.r-project.org/doc/manuals/r-release/R-intro.pdf>.

Wickham, H. (2010). *A Layered Grammar of Graphics*. Journal of Computational and Graphical Statistics, 19:1, 3-28, DOI: 10.1198/jcgs.2009.07098.

Wickham, H. (2014). Tidy Data. *Journal of Statistical Software*, 59(10), 1-23.

Wickham, H., and Grolemund, G. (2017). *R for Data Science*. <http://r4ds.had.co.nz/>.

Wickham, H., Chang, W., Henry, L., Pedersen, T., Takahashi, K., Wilke, C., Woo, K. and Yutani, H. (2019). *Ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics*. <https://CRAN.R-project.org/package=ggplot2>.

Wilkinson, L. (2005). *The Grammar of Graphics (Statistics and Computing)*. Secaucus, NJ, USA: Springer-Verlag New York, Inc.

Xie, Y. (2017). *Dynamic Documents with R and knitr*. Boca Raton, Florida: CRC Press.

Xie, Y., Allaire, J. and Grolemund, G. (2018). *R markdown: The definitive guide*. CRC Press. <https://bookdown.org/yihui/rmarkdown/>.

Xie, Y. (2019). *xaringan: Presentation Ninja*. <https://CRAN.R-project.org/package=xaringan>.

Xie, Y., Cheng, J., and Tan, X. (2019). *DT: A Wrapper of the JavaScript Library ‘DataTables’* <https://CRAN.R-project.org/package=DT>.

1. Shiny is an R package that provides a framework for building web applications using R. <http://shiny.rstudio.com/articles/interactive-docs.html> [↑](#footnote-ref-1)