DATA SCIENCE PROJECT

1   Overview

We have chosen to use the data about the Game Of Thrones series. Several plots, charts and maps have been implemented in a Shiny application, with dplyr, ggplot, sf and cartography.

To organize them, we have created a page with a top level navigation bar (navbarPage object).
In this bar, we have used several embedded menus within the navbar (navbarMenu objects). Each menu contains several tab panels (tabPanel objects).

The Deaths menu proposes the plots related to the deaths in the serie :
   ● Deaths by location
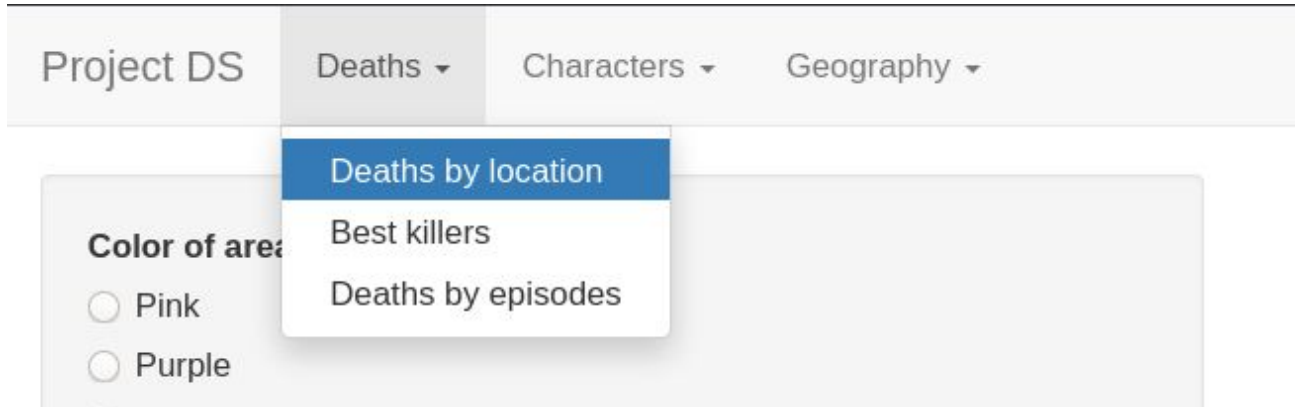   ● Best killers
   ● Deaths by episodes
The Characters menu proposes the plots related to the characters :
   ● Characters info
   ● Distributions
The Geography menu proposes the plots related to the geography of the GoT world :
   ● Demography
   ● Topology
   ● Landscapes

The following figure shows how the navigation bar works.

All the charts have dynamic sizes, using the session data to rescale the height to auto when the size of the window changes.

2   Deaths plots

2.1   Deaths by location

This chart displays the map of the GoT world and the number of deaths in the different locations as areas. Of course, the size of areas is proportional to the number of deaths.

In the Shiny application, you may choose a different color for the location areas by using the radio buttons and you may specify the range of number of deaths that must be displayed by playing with the slider.

2.2   Best killers

This chart displays the best killers in the series. A bar chart is used to display the number of victims of each killer. You can also see the gender repartition in the victims of the killers.

In the Shiny application, you can select how many killers you want to display in the chart.

2.3   Deaths by episodes

This chart displays, season by season, the number of deaths for each episode of the season. One chart is displayed per season and the charts are organized in a facet_wrap with 2 rows.

In the Shiny application, you can select the seasons you want to be displayed by using the check box.

3   Characters plots

3.1   Characters info

Here, there are 3 dropdown lists.

The first one, allows you to choose between 4 datasets (scenes, appearances, characters and episodes). Then the summary of the selected dataset will be displayed on the right.

Then, we can choose a character in the next dropdown list, and a graph will be displayed depending on the last list selected option.

The first option, "Episodes by season" will display the number of episodes the selected character appears on, for each season, using a bar chart.

The second one, "Time screen", represents the evolution of the time screen appearance, for the selected character, for each episode he appears on. It uses a line chart with dots to highlight the episode where the character appears.

3.2   Distribution

This section represents some distributions for the GoT dataset, these "no-input" functions can be reached each by its specified radio button.

1. Gender distribution per house.
This bar chart displays the proportions of females and males for each house of Got.

2. Survivals proportion per season.
This pie chart displays the proportion of survivors of GoT, collected by each episode and grouped by season number.

4   Geography plots

4.1   Demography

The demography of the GoT world is displayed by either location density or location size. A slider, from 1 to 5, enables to choose the range of the location sizes taken into account.

1. The location density map uses stat_density2d with the (x,y) coordinates of the centroids of the locations, with each row replicated as many times as the size of the location, to produce a density graph weighted by location size on the GoT map. The theme of the cartography package has been reproduced in ggplot to have a more harmonious rendering with the other maps.

2. The location size map uses the cartography package to display the locations by size.

4.2   Topology

A zoom is done on each continent. The selected continent is passed as input with the radio button, and a detailed map of the continent is displayed with the cartography package. Here, all the landscapes are displayed: forests, mountains, steppes, swamps, islands, lakes, rivers, roads and the wall. The major locations are also displayed using an st_contains to compare the coordinates of the locations with the area of the selected continent.

4.3   Landscapes

Landscape statistics are displayed in charts by either surface area or landmass.

1. The landscapes by surface area displays the area of each type of landscape (forests, mountains, steppes, swamps, islands, lakes, rivers, roads and the wall) by surface area and by continent in a facet_grid. The labels are in scientific notation with 2 significant digits, and the x axis is logarithmic. An average of the width has been done with data found online for one dimensional landscapes to convert them to areas (135.4 m for rivers, 3.5 m for roads and 91 m for the wall).

2. The landscapes by landmass displays the total landmass by continent in a pie chart.