Lab 03: Univariate Plots and Maps

**Handout date:** Wednesday, September 18, 2019

**Due date:** Wednesday, September 25, 2019 at the beginning of class as hardcopy

*This lab counts 4 % toward your total grade*

**Objectives:** In this lab   
[a] explore the library **RColorBrewer** and generate maps with 

[b] you will practice generating meaningful statistical graph from the family of bar-plots and histograms.  
Note: Box-plots will be exercised among other topics in Lab04.

**Format of answer:** Your answers (graphs and verbal description) should be handed in as ***hard-copy*** in ***one*** document. Add a running title into the header of the document with the following information: ***your name***, ***Lab03*** and ***page numbers***. Label each answer properly starting with its task number. Maintain the sequence of questions. Format any code and computer output properly before inserting it into the document with your answer. -code and text output need to be in a ***monospaced*** font (i.e., fixed-pitch font) such as Courier New so proper spacing and alignments are preserved. Excessive, but irrelevant, output will lead to a deduction of your accumulated points.

## Task 1: Color Brewer (0.6 points)

Link the library **RColorBrewer** to your  session and ***explore the online help*** and the function **display.brewer.all( )**.

The table below shows three different color palettes with 7 classes, which are tailored toward colorblind readers.

|  |  |  |
| --- | --- | --- |
| *Palette 1* | *Palette 2* | *Palette 3* |

[a] For each palette identify its “name” in the library **RColorBrewer** and underlying type, i.e., *sequential*, *diverging* and *qualitative*. Justify your answer. (0.3 points)

Palette 1:

Palette 2:

Palette 3:

[b] Recreate the three palettes with ***9 classes*** and put their properly sized images into the table below. (0.3 points)

|  |  |  |
| --- | --- | --- |
| *Palette 1* | *Palette 2* | *Palette 3* |

## Task 2: Mapping (1.2 points)

You are given three **Esri** shape files. These files are packed into the zipped file **Italy.zip**. All geographies are in the ***lat***/***long*** format and therefore not projected. Note:  takes care of projecting these files properly.

The maps you will generate consist of two area layers: [a] the countries neighboring Italy (layer **Neighbors.shp**) and [b] added on top of it the 95 provinces of Italy (layer **Provinces.shp**). Make sure that the ***frame of the plot window*** is sized properly to embed Italy. The base map below shows these two layers in a properly sized window frame.

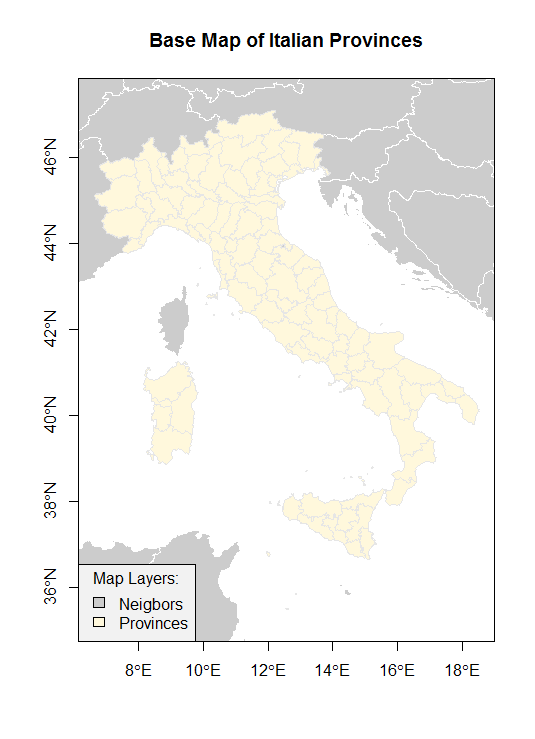


Figure 1: Base Map of the Italian Provinces

You will generate three color maps displaying different map themes. For your lab answers please show these maps in ***color***. Each map should be properly ***framed***, have a proper ***title*** and ***legend***, show the ***neighboring countries*** as spatial reference frame. You can use the mapping functions in the package **DallasTracts**. Just show the code for your maps.

[a] Generate a map showing the Italian regions, which are stored in the variable **REGION**. Show the relevant code used to generate the map. (0.3 points)

[b] Generate a map showing the total fertility rate (number of births per woman) in the Italian provinces using 7 classes, which is stored in the variable **TOTFERTRAT**. Show the relevant code used to generate the map. (0.3 points)

[c] Generate a map showing the gender ratio in the 95 provinces, which can be calculated with the transformation **logMigRatio <- log(*shp*$INFLOW/*shp*$OUTFLOW)**, where ***shp*** refers to the name of your imported shape-file. What is the neutral break-point for the variable **logMigRatio**. Use in total 8 classes but select the appropriate number of classes for the below and above breakpoint branches of the underlying distribution of **genderRatio**. Justify your choice. Show the relevant code used to generate the map. (0.6 points)

## Study Kabacoff’s Chapter 6 “Basic Graphs” pp 117-136

In order to solve the remaining tasks, you need to study Kabacoff’s Chapter 6 “Basic Graphs” pp 117-136 and adjust the code examples given there to generate the requested graphs for the different datasets.

## Task 3: Bar-plots and their Derivatives (1.3 points)

To generate bar-plots and their derivatives you will be working with data from the Titanic disaster on April 16, 1912, in the North Atlantic.

The screenshot below shows some code, which was used to prepare the data to generate the plots in this task. In order to generate the plots in tasks 3 [e]-[g], you need to enter and execute this code.

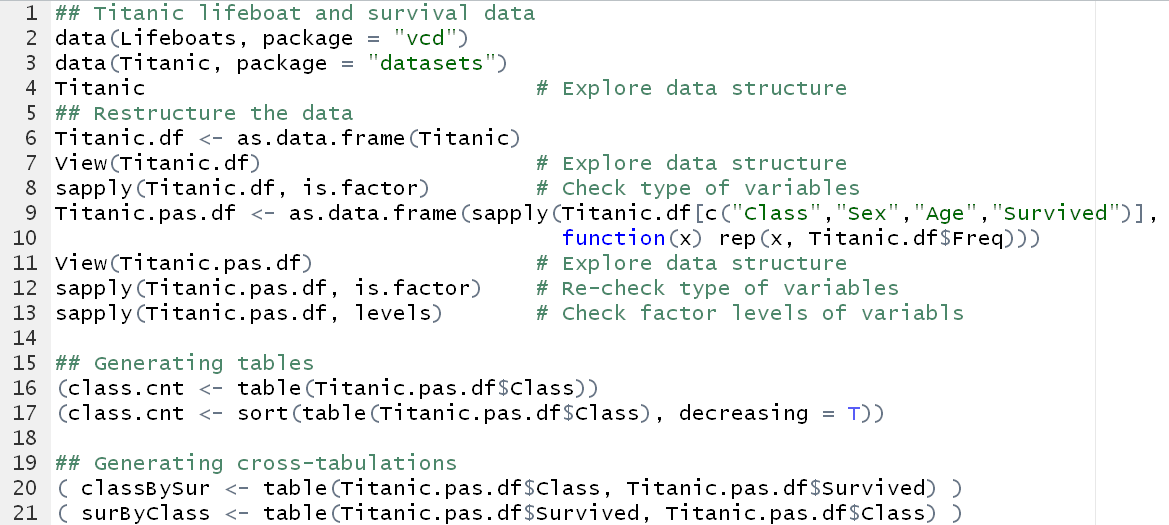


Figure 2: Preparation of the data for histogram tasks

[a] Discuss how the organization of the **Titanic** data in line 4 changes for the **Titanic.df** data-frame in line 7 to the **Titanic.pas.df** data-frame in line 11. (0.3 points)

[b] What are the **sapply( )** function calls in lines 8, 12 and 13 doing? (0.2 points)

[c] How do the two tables in lines 16 and 17 differ? (0.1 points)

[d] What is the difference between the two cross-tabulations in lines 20 and 21? (0.1 points)

[e] Recreate the bar-plot shown below and show the code that you have used to generate the plot. Which of the tables in lines 16 and 17 should you use? (0.2 points)

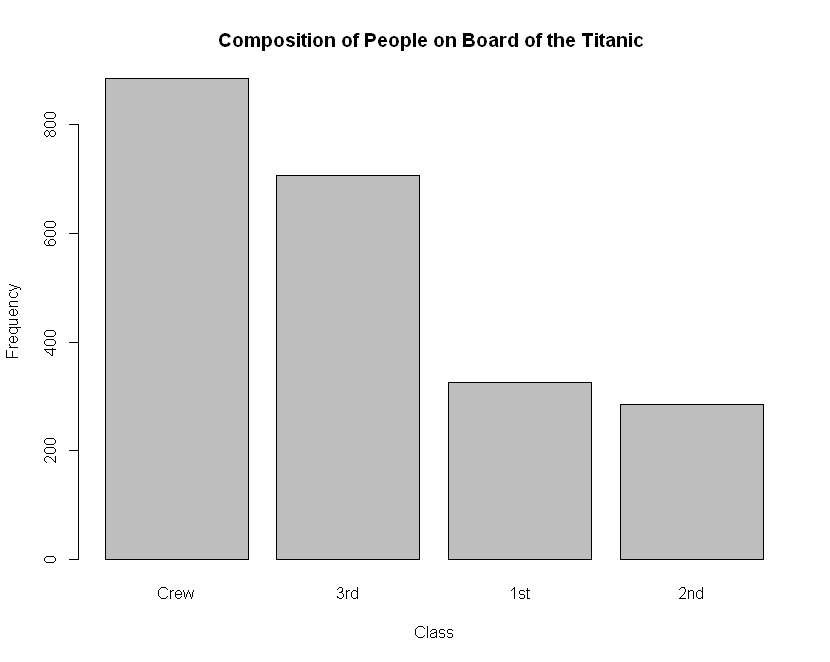


Figure 3: Histogram sorted by frequency

[f] Recreate both bar-plots shown below and show the specific code that you have used to generate these plots. Briefly discuss what the difference between both plots is. Hint: look at the scales for the y-axes. (0.4 points)

|  |  |
| --- | --- |
| Figure 4: Stacked bar-plot | Figure 5: Spinogram |

## Task 4: Histograms with Kernel density overlays (0.9 points)

Open the old faithful geyser dataset with **data("faithful", package="datasets")**. For information on this dataset consult the online help.

Generate for the variable **eruptions** time a ***density histogram*** with observation ticks at the bottom of the x-axis and an overlaid ***density curve***. The  code generating a plot is:

**lines(density( ))**). The  code generating a plot is:

**hist(faithful$eruptions,**

**breaks = seq(1.5,5.3,by=*selBy*), probability=TRUE,**

**main="Eruption Times of Old Faithful")**

**rug(jitter(faithful$eruptions))**

**lines(density(faithful$eruptions, bw=*selBw*), col="red")**

[a] Explain what the **breaks**-option of the **hist( )** function is defining. (0.1 points)

[b] Explain why the option **probability=TRUE** needs to be set to true in order for the kernel density curve to display properly on top of the histogram. (0.1 points)

[c] Explain what the option **bw** in the **density( )** function is doing. See for instance, <https://en.wikipedia.org/wiki/Kernel_density_estimation> or BBR pp 410-415. (0.1 points)

[d] Generate and show the histogram-density plot with the parameters **selBy=0.05** and **selBw=0.02**. What is wrong with the selected set of parameters? (0.2 points)

[e] Generate and show the histogram-density plot with the parameters **selBy=0.475** and **selBw=0.8**. What is wrong with the selected set of parameters? (0.2 points)

[f] Find a set of parameters **selBy** and **selBw** by experimentation that informatively displays the underlying distribution of the erruption times. Show your final histogram-density plot and ***justify*** your selection of parameters. (0.2 points)