Lab 03: Univariate Plots and Maps

**Handout date:** Wednesday, September 16, 2020

**Due date:** Wednesday, September 23, 2020 by midnight as Word document into eLearning’s SubmitLab03 link

*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.   
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*** (note palette 1 only allows for 8 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.

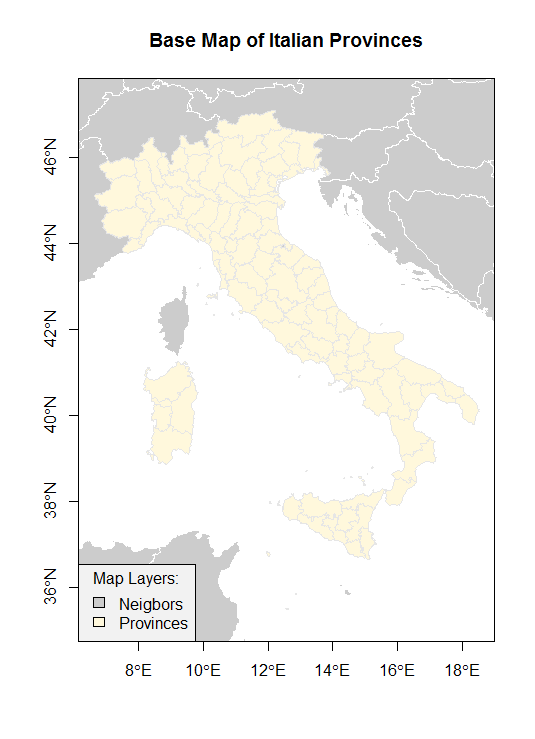


Figure 1: Base Map of the Italian Provinces

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 of zero of the underlying distribution of **logMigRatio**. Justify your choice. Show the relevant code used to generate the map. (0.6 points)

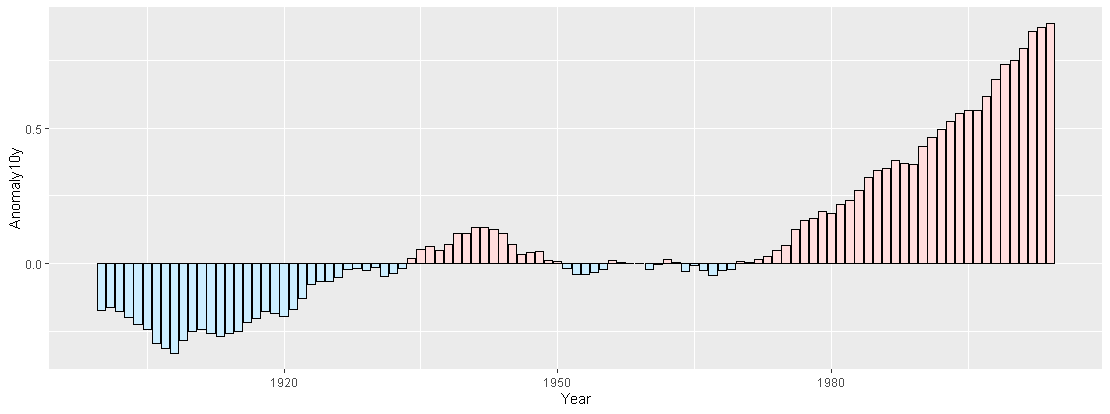
## Task 3: Statistical Graphs (1.2 points)

The following graphs are reproductions from the [R Graphics Cookbook, 2e](https://r-graphics.org/index.html). One of the objectives of this task is that you visit the **R Graphics Cookbook** and explore the graphs on display there.

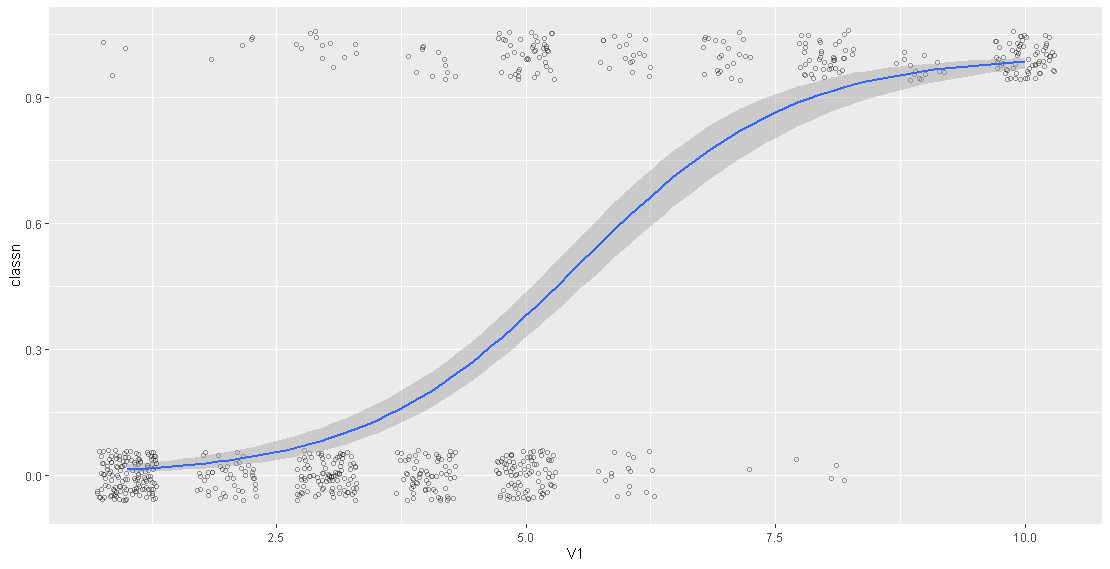
Study the sections associated with the graphs, which are displayed below. Its sections document how these graphs were built.

You job in this is to *reproduce* these graphs and *show the code*, which generated them. The data used in the **R Graphics Cookbook** are available in **library("gcookbook")**.

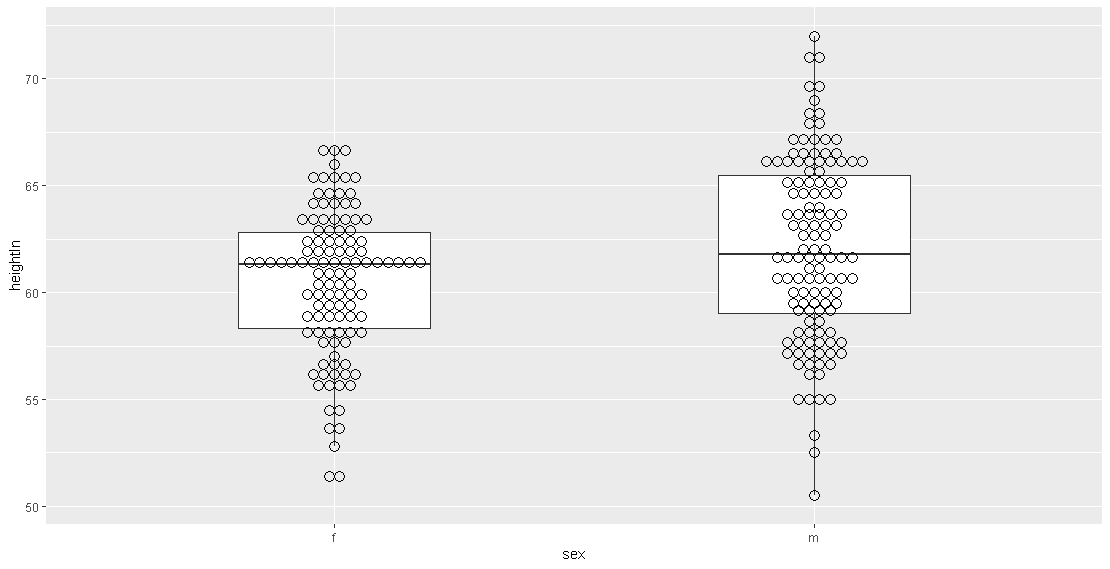
[a] Reproduce the graph below and show the code (0.3 pts)



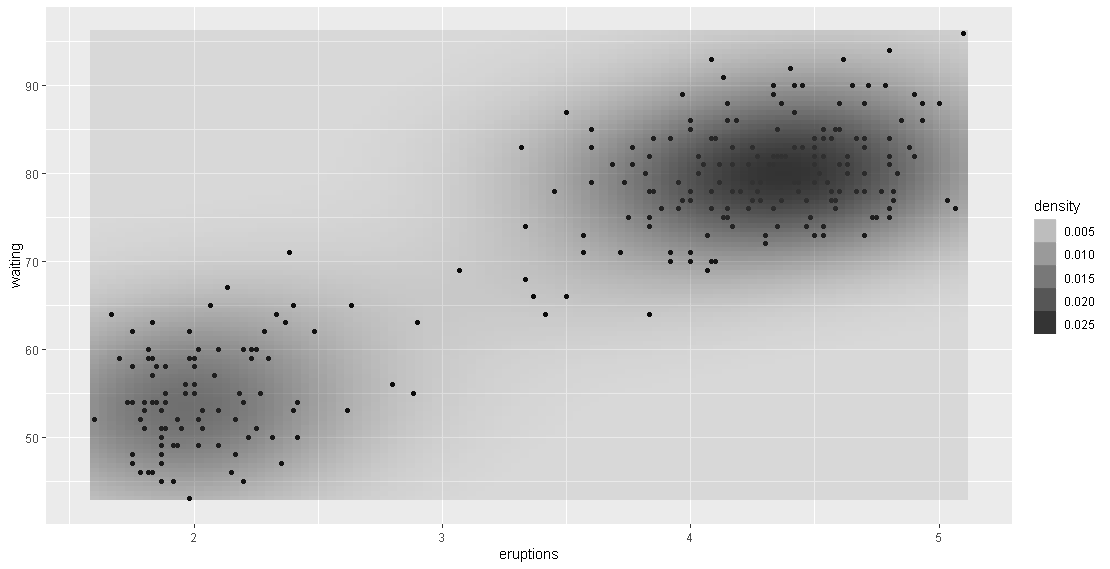
[b] Reproduce the graph below and show the code (0.3 pts)



[c] Reproduce the graph below and show the code (0.3 pts)



[d] Reproduce the graph below and show the code (0.3 pts)



Hint: For some graphs you would also need to look at the data transformation that have been applied before the graph is rendered.

## Task 4: Histograms with Kernel density overlays (1 point)

Open the **Concord1.sav** file and generate three histograms with the code

**hist(concord$income, breaks=seq(0,100,by=???), freq=FALSE,**

**ylim=c(0,0.05), xlab="Income in $1000",**

**main="Income Distribution in the City of Concord")**

**rug(jitter(concord$income, factor=1), side=1)**

**lines(density(concord$income, bw=???), lwd=2)**

a. Show the histogram with the associated density curve for **by=2.5** and **bw=1**.

b. Show the histogram with the associated density curve for **by=5** and **bw=2.5**.

c. Show the histogram with the associated density curve for **by=20** and **bw=10**.

d. Justify, which of the three parameter combinations most realistically reflects the distribution of the income.

e. Show the ***stem-and-leaves*** plot (function **stem( )** )to identify whether the income variable has the tendency to cluster around particular values.

Hint: The income was reported by sampled household heads. Do people have the tendency to round their responses?