Online Academic Data Analysis Bootcamp Using Open-Access Program R: Essentials. Session 4, Base Plots in R

Patrick Njage

Aug 17th 2023

# About the data

The General Social Survey is a high-quality survey which gathers data on American society and opinions, conducted since 1972. This data set is a sample of 500 entries from the GSS, spanning years 1973-2018, including demographic markers and some economic variables. Note that this data is included for demonstration only, and should not be assumed to provide accurate estimates relating to the GSS.

# Preliminary

Read the GSSsubset into R

To avoid “figure margins too large error” first change graphical parameters by running the following code

*par(mar=c(1,1,1,1))*

# Exercise 1

1. Create a histogram of income changing the breakpoints between histogram bars to 25.
2. Create a density plot of income and fill the density plot with a colour of choice.
3. Create a barplot of the degrees variable giving each bar a different colour. Hint: first tabulate the number of participants for each degree level using *table(GSSsubset$degree)* and allocate this an object name to be used for plotting.
4. Change the barplot in d) above into a horizontal bar plot (adding the argument *horiz = TRUE*).
5. Save the barplot using:

* RStudio Plots Panel, and
* directly as a pdf to your working directory.

# Exercise 2

1. Plot a pie chart of the degree variable. Hint: use table function to create table as done in Exercise 1 above about change this table into a dataframe using *as.data.frame(table)* then asign a name to this dataframe.
2. Convert this pie chart into a 3D pie chart.

# Exercise 3

1. Make a **multiple groups** boxplot of income classified into sex and degree giving different colors for each group.
2. Provide a relevant title and axis labels and remove frame.
3. Save this a pdf of a suitable page size in your working directory. Hint: PDF’s are 7x7 inches by default, and each new plot is on a new page. The page size can be changed to for instance 20x15 inches by adding the arguments *pdf(“plots.pdf”, width=20, height=15)*. Change the page size to save into a suitably presentable plot.

# Exercise 4

1. Plot the bivariate relationship relationships between hours of work (hrswrk) and income:

* provide title and axis labels, and
* Add a regression and a lowess line.

1. Use function scatterplot3D to make a 3D plot similar to a) above.

* Change color of the boxes grouping by degree.
* Add grids and remove the box around the plot.
* Change axis label names.

# Exercise 5: Graphical Parameters

1. Randomly select 200 rows from the dataset GSSsubset to create GSSsubset2.
2. Make a plot of hours of work (hrswrk) (y-axis) versus income (x-axis) using GSSsubset2.
3. Add a main title and x and y axis labels using a separate title() code line.
4. Assuming some cut-off levels of interest for income and hours of work are known from literature. Add reference lines to the graph using the abline( ) function:

* solid horizontal lines at y= 20 and 55,
* dashed blue verical lines at x = 10000,50000 and 100000.

# Exercise 6

Multiple plots can be combined into one overall graph par( ) function.

1. Create the following four figures from the GSSsubset2 arranged in 2 rows:

* scatter plot of plot income versus hours of work (insert title),
* scatterplot plot age verus hours of work (insert title),
* Make a histogram of height (insert title),
* Make a boxplot of income by degree, add title, y and x axis labels, and different colour for each degree.

1. Save the graph from a) above directly to your working directory in a suitable page size.