Day 1 - Exercise 3
Getting started with ggplot

Data sets

Fuel economy data 'mpg'

The data set 'mpg' comes with the *ggplot2* package and contains fuel economy data from 1999 to 2008 for 38 popular models of cars. A data frame with 234 rows and 11 variables. Be aware that 'mpg' has a *tibble* structure (subclass of data.frame).

Pharmacokinetics of Theophylline 'Theoph'

The 'Theoph' data frame has 132 rows and 5 columns of data from an experiment on the pharmacokinetics of theophylline. There are 12 subjects with repeated measurement of theophylline concentration (11x) over time.

Famous iris data 'iris'

This famous (Fisher's or Anderson's) 'iris' data set gives the measurements in centimeters of the variables sepal length and width and petal length and width, respectively, for 50 flowers from each of 3 species of iris.

Get started

Task 0 - 'prepared' R file

- Download Rfile *day1_ex3_ggplot_vYYYYMMDD.R* from GitHub into your folder 'R/Rfiles'
- Open the file in RStudio
- Use 'Outline' in RStudio to get an overview of the R file structure
- Run the code till the heading 'Task 1 - Barcharts'
- Have look at the three datastes which were prepared

Different types of plots using dataset 'mpg'

Task 1 - Bar charts

The bar chart illustrates the frequency of manufacturers. Run the code for the bar chart and then adapt it.

a) change color and fill of bars

- Hint: within 'geom_bar' try something like: fill = "darkgreen"

b) change background of graph using 'theme_bw()'

- Hint: look at documentation of theme_bw() - are there other interesting themes?

c) we need to change the angle of the labels on the x-axis

- Hint: search in a search engine with following key words: ggplot change angle label x axis
- try different values for angle (e.g., 45)
- look at documentation of element_text()
  - maybe we need to work with options hjust (and vjust?)

d) Use the same plot and change from absolute frequencies to relative frequencies.

- Hint: write 'aes(y = after_stat(prop), group = 1)' in geom_bar()

e) Do you find a way to change the y axis to display %?

- Hint: look at documentation of scale_y_continuous()

f) Can you get rid of the 'prop' as y axis title?

Task 2 - Histogram

The histogram illustrates the engine displacement in litres (displ). Run the code for the histogram and then adapt it.

a) change color and fill of bars
b) change background of graph using 'theme_bw()' (or one you prefer)
c) we might want to change the interval width into 0.25 liter

- Hint: look for 'binwidth' in documentation of geom_histogram()

d) change x axis title to 'Engine displacement [l]' using xlab()

- Hint: search in a search engine with following key words: ggplot change angle label x axis

e) save the output in object p_hist_displ

- Hint: 'p_hist_displ <-'

Note: The 'p_hist_displ <-' saves the ggplot on the variable name 'p_hist_displ' and therefore we can refer later back to this plot.

Histogram for each category

f) Reuse the histogram code from above and retrieve histograms by "type" of car (class)

- Hint: facet_grid()

Task 3 - Box plot

The box plot illustrates the engine displacement in litres (displ). Run the code for the box plot and then adapt it.

a) Adapt the box plot such that it illustrates the engine displacement in litres (displ) by "type" of car (class)

- Hint: we want to have the "type" of car (class) on the x axis, i.e. aes() needs to be adapted

Task 4 - Combine different plots in one big plot

a) We need to install the package 'cowplot' in order to use the function plot_grid()

- Tools → Install packages . . .
- type 'cowplot'
- press 'Install'
- load the library 'cowplot'

b) Adapt the given code to

- add panel labels
- have the panels underneath each other

Note: The use of package name together with '::' allows to

- use a function from a package without loading the package;
- to define which function we want to call in case two/several packages defined a function with the exact same name

Task 5 - Scatter plot

a) Create a scatter plot with engine displacement (displ) on x axis and highway miles per gallon (hwy) on y axis.

- Hint: use geom_point()

b) Adapt axis titles accordingly

c) As highway miles per gallon (hwy) is only measured in integers it is likely that we have overlapping observations

- Hint: use geom_point(alpha = 0.3) and look into documentation for the meaning of 'alpha'

d) We want to differentiate if number of cylinders (cyl) is >5 or not

- Hint: add 'color = (cyl > 5)' to the aes() statement

e) Adapt the legend title to 'Number of cylinders'

- Hints: use labs(color = "ABC")

f) Adapt legend from TRUE/FALSE to >5 and <=5 and the colors used to 'blue' and 'red'

- Hint: scale_color_manual() and within that labels and values

Different types of plots using 'Theop'

Task 6 - Line plot

a) Create a plot with points and lines to show theophylline concentration in the subjects over time.

- Hint: you need to state in aes() that the data is grouped by the patients, i.e. 'group = Subject'
- Hint: use the geometric objects 'geom_point()' and 'geom_line()'

b) Adapt axis titles accordingly

c) We want to differentiate if dose of theophylline administered orally to the subject (Dose) is >4.5 or not

- Hint: add 'color = (cyl > 5)' to the aes() statement
- use your own colors

d) Adapt the legend title and legend accordingly

Plots using 'iris'

Task 7 - iris

a) Create a scatter plot as in the slide 'Example - Iris: adding title'
b) Create a box plot as in the slide 'Example - Iris: using another geom'
c) Combine these two plots within one plot as panel A and panel B