---

title: "Lab 3 solutions"

author: "Ines Arana"

date: "2022/23"

output: html\_document

---

```{r setup, include=FALSE}

knitr::opts\_chunk$set(echo = TRUE)

```

Clearing the environment.

```{r}

rm(list=ls())

```

Loading library (package) tidyverse. It loads the ggplot2 library used to generate the plots as well as other libraries.

```{r warning=F}

library(tidyverse)

```

Loading the data file as a data frame.

````{r}

consumptionNoOil <- read.csv("consumptionNoOil.csv", header=T, stringsAsFactors=T)

summary(consumptionNoOil)

head(consumptionNoOil, 20)

```

Alternative, loading the data file as a tibble. Use either the above (as a dataframe) or the below (as a tibble) but not both.

```{r}

consumptionNoOil <- read\_csv("consumptionNoOil.csv")

consumptionNoOil <- consumptionNoOil |> mutate(across(where(is.character),as\_factor))

summary(consumptionNoOil)

head(consumptionNoOil, 20)

```

First plot: a scatterplot using Year in the x axis and MTOE in the y axis.

```{r}

p <- ggplot(data= consumptionNoOil, aes(x=MTOE,y=EstMTOE))

p <- p + geom\_point()

p

```

Adding some colour - green points.

```{r}

p <- ggplot(data= consumptionNoOil, aes(x=MTOE,y=EstMTOE))

p <- p + geom\_point(colour="darkgreen")

p

```

Adding colour to indicate the source of energy (oil, gas or renewables).

```{r}

p <- ggplot(consumptionNoOil, aes(MTOE, EstMTOE, colour=Source))

p <- p + geom\_point()

p

```

In the above plot it is easy to see that coal-related points generally have low MTOE and EstMTOE values when compared to gas and renewables data points.

Now we try a different point shape depending on the Source of energy value.

```{r}

p <- ggplot(consumptionNoOil, aes(MTOE, EstMTOE, shape=Source))

p <- p + geom\_point()

p

```

We can combine shape and colour to illustrate the source of energy.

```{r}

p <- ggplot(consumptionNoOil, aes(MTOE, EstMTOE, colour = Source, shape=Source))

p <- p + geom\_point()

p

```

Using a line. This is clearly \*\*not a good idea\*\*, but it illustrates why there is a need to separate the data according to the value of the Source of energy column.

```{r}

p <- ggplot(consumptionNoOil, aes(Year, MTOE))

p <- p + geom\_line()

p

```

More appropriate use of lines, one line per Source value.

```{r}

p <- ggplot(consumptionNoOil, aes(Year, MTOE, colour=Source))

p <- p + geom\_line()

p

```

An alternative is to group points by Source value. We get 3 lines, each representing a different source value.However, which line represents which source value is not conveyed.

```{r}

p <- ggplot(consumptionNoOil, aes(Year, MTOE, group = Source))

p <- p + geom\_line()

p

```

Changing the colour and thickness of the lines.

```{r}

p <- ggplot(consumptionNoOil, aes(Year, MTOE, group = Source))

p <- p + geom\_line(size = 2, colour="darkgreen")

p

```

In the above, we can see how related data is connected via the line. Below we can see that if we just use points, that connection is not visible. We can see 3 points per year,but we do not know how they relate to previous/future years.

```{r}

p <- ggplot(consumptionNoOil, aes(Year, MTOE, group = Source))

p <- p + geom\_point()

p

```

The plot below shows how several geoms can be added to a plot. In this case we have points and lines. The source value for each line is unknown.

```{r}

p <- ggplot(consumptionNoOil, aes(Year, MTOE, group = Source))

p <- p + geom\_point()

p <- p + geom\_line()

p

```

Below we have a similar plot, but the points have been modified to light blue triangles with red lines.

```{r}

p <- ggplot(consumptionNoOil, aes(Year, MTOE, group = Source))

p <- p + geom\_line()

p <- p + geom\_point(shape = 24, colour = "red", size = 4, fill = "lightblue")

p

```

Different groups on different layers. One smooth for all groups

```{r}

p <- ggplot(consumptionNoOil, aes(Year, MTOE, group = Source))

p <- p + geom\_line()

p <- p + geom\_smooth(aes(group = 1), formula='y~x',method = "lm", size = 2, se = F)

p

```

Compare the above with the below, where we get one linear model per source value.

```{r}

p <- ggplot(consumptionNoOil, aes(Year, MTOE, group = Source))

p <- p + geom\_line()

p <- p + geom\_smooth(aes(group = Source), formula='y~x',method = "lm", size = 1, se = F)

p

```

Creating a histogram with the pressure1 dataset.

```{r}

pressure1Data <- read.csv("pressure1.csv", header=T)

p <- ggplot(pressure1Data, aes(Pressure))

p <- p + geom\_histogram(colour="blue", fill="green", binwidth=5)

p

```

### Exercise 1

Loading the dataset.

```{r}

ProvenOilReservesZerosT <- read.csv("ProvenOilReservesZerosT.csv", header=T, stringsAsFactors=T)

```

Note that in this dataset the year column (first column) is called MTBarrels, to indicate that the unit of measure for the dataset is million ton barrels.

The dataset contains '-'. In R, we would normally have NA (not available) or zero.

Replacing '-' by NA.

```{r}

ProvenOilReservesZerosT |> mutate\_all(funs(replace(., .== "-", as.double(NA))))

```

### Exercise 2

```{r}

p <- ggplot(ProvenOilReservesZerosT, aes(MTBarrels, Denmark))

p <- p + geom\_point()

p

```

### Exercise 3

```{r}

p <- ggplot(ProvenOilReservesZerosT, aes(MTBarrels, Denmark))

p <- p + geom\_point(shape=22, colour="darkblue", size=4)

p

```

#### Exercise 3 - if we want the squares filled in dark blue

```{r}

p <- ggplot(ProvenOilReservesZerosT, aes(MTBarrels, Denmark))

p <- p + geom\_point(shape=22, colour="darkblue", size=4, fill="darkblue")

p

```

### Exercise 4

```{r}

p <- ggplot(ProvenOilReservesZerosT, aes(MTBarrels, Italy))

p <- p + geom\_line(colour="orange", size=2)

p

```

### Exercise 5

Note that \*\*using colour to emphasize size may not be the best option\*\*.

```{r}

p <- ggplot(ProvenOilReservesZerosT, aes(MTBarrels, United.Kingdom, colour=United.Kingdom))

p <- p + geom\_point()

p

# Alternative using the factor function

p <- ggplot(ProvenOilReservesZerosT, aes(MTBarrels, United.Kingdom))

p <- p + geom\_point(aes(colour=United.Kingdom))

p

```

### Exercise 6

```{r}

p <- ggplot(ProvenOilReservesZerosT, aes(MTBarrels, Norway, size=Norway))

p <- p + geom\_point(shape=24, colour="red")

p

```

# Exercise 6 if we want the triangles filled in red

```{r}

p <- ggplot(ProvenOilReservesZerosT, aes(MTBarrels, Norway, size=Norway))

p <- p + geom\_point(shape=24, colour="red", fill="red")

p

# Alternative

p <- ggplot(ProvenOilReservesZerosT, aes(MTBarrels, Norway))

p <- p + geom\_point(aes(size=Norway), shape=24, colour="red", fill="red")

p

```

### Exercise 7

Loading a second a dataset

```{r}

ProvenOilReserveWEurope <- read.csv("ProvenOilReserveWEurope.csv", header=T, stringsAsFactors=T)

```

### Exercise 8

```{r}

p <- ggplot(ProvenOilReserveWEurope, aes(Year, MT.Barrels, group=Country))

p <- p + geom\_line()

p

```

### Exercise 9

```{r}

p <- ggplot(ProvenOilReserveWEurope, aes(Year, MT.Barrels, group=Country))

p <- p + geom\_line(colour="purple")

p

```

### Exercise 10

```{r}

p <- ggplot(ProvenOilReserveWEurope, aes(Year, MT.Barrels, group=Country, colour=Country))

p <- p + geom\_line()

p

```

### Exercise 11

```{r}

p <- ggplot(ProvenOilReserveWEurope, aes(Year, MT.Barrels, group=Country))

p <- p + geom\_line()

p <- p + geom\_smooth(aes(group = 1), formula = y~x, method = "lm", size = 2, se = F)

p

```

### Exercise 12

```{r}

p <- ggplot(ProvenOilReserveWEurope, aes(Year, MT.Barrels, group=Country, colour=Country))

p <- p + geom\_line()

p <- p + geom\_smooth(aes(group = 1), formula = y~x, method = "lm", size = 2, se = F)

p

```

### Exercise 13

```{r}

p <- ggplot(ProvenOilReserveWEurope, aes(Year, MT.Barrels, group=Country, colour=Country))

p <- p + geom\_point()

p <- p + geom\_smooth(aes(group = 1), formula = y~x, method = "auto", se = F)

p

```

### Exercise 14 - histogram

We choose a bindwidth of 3, a red line (colour) and orange fill.

```{r}

pressure1Data <- read.csv("pressure1.csv", header=T)

p <- ggplot(pressure1Data, aes(Pressure))

p <- p + geom\_histogram(colour="red", fill="orange", binwidth=3)

p

```

Note how a smaller binwidth leads to more (but narrower) "bars".

We now choose a bindwidth of 8, a darkgreen line (colour) and yellow fill.

```{r}

p <- ggplot(pressure1Data, aes(Pressure))

p <- p + geom\_histogram(colour="darkgreen", fill="yellow", binwidth=8)

p

```

Note there are less bars now, but these are much wider.

### Heat Map

```{r}

# theme\_set(theme\_classic())

between90and01 <- ProvenOilReserveWEurope |> filter(Year %in% c(1990:2001))

```

```{r}

#p <- ggplot(ProvenOilReserveWEurope, aes(Year, Country))

p <- ggplot(between90and01, aes(Year, Country))

p <- p+ geom\_tile(aes(alpha=MT.Barrels), fill= "red")

p

```

```{r}

## generates all points. From http://www.sthda.com/english/wiki/r-plot-pch-symbols-the-different-point-shapes-available-in-r

generateRPointShapes<-function(){

oldPar<-par()

par(font=5, mar=c(0.5,0,0,0))

y=rev(c(rep(1,6),rep(2,5), rep(3,5), rep(4,5), rep(5,5)))

x=c(rep(1:5,5),6)

plot(x, y, pch = 0:25, cex=1.5, ylim=c(1,5.5), xlim=c(1,6.5),

axes=FALSE, xlab="", ylab="", bg="red")

text(x, y, labels=0:25, pos=3)

par(mar=oldPar$mar,font=oldPar$font )

}

p <- generateRPointShapes()

```