What a beauty!

Claudius

## 2022-04-06

Packages used

```
library(tidyverse)
library(DataScienceExercises)
library(knitr)
```

Exploring flight data

In this short text we explore the following data set on flights departing from New York.

arr_delay	dep_delay	month	carrier	distance
-39	-4	4	DL	2248
-22	-4	12	AA	1389
0	-4	1	B6	1076
-8	-1	7	UA	1608
-7	-4	3	DL	1035

To have a first look on the relationship of the variables, consider the following scatter plots:
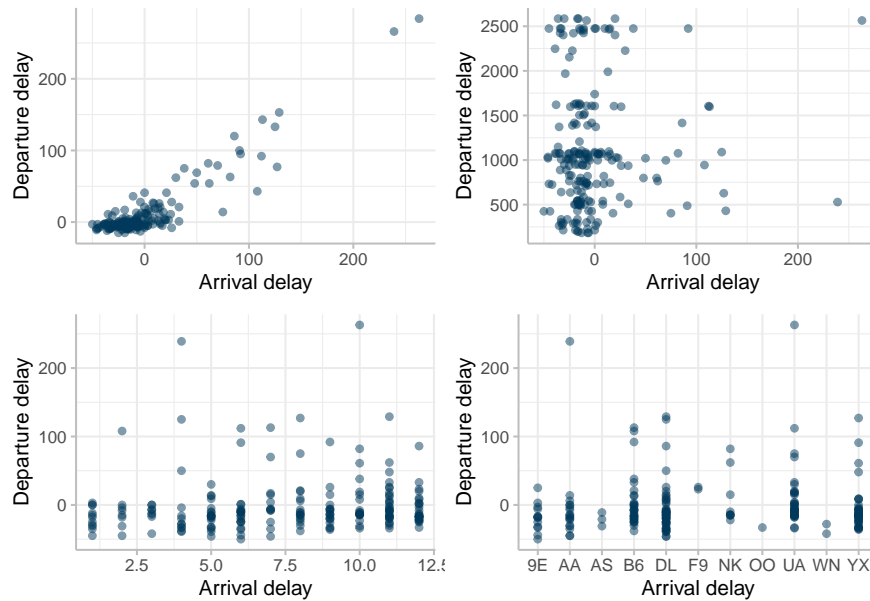
These plots suggests that there is a strong correlation between departure and arrival delay. To compute the correlation we might use the following R code:

```
cor_coef <- cor(base_data$arr_delay, base_data$dep_delay)
```

This produces a correlation coefficient of 0.911, suggesting that there is indeed a very strong correlation.

But is it significant? Lets check it using the Pearson correlation test:

```
c_test <- cor.test(
  x = base_data$arr_delay,
  y = base_data$dep_delay,
  method = "pearson")
```

The most relevant statistics are:

t-stat	df	p-val	95% conf interval
31.16616	198	0	[0.885; 0.932]

Of course, these are just preliminary results, from a methodological point of view there is still much to do. . .

The corrections we did

To make this document look *much* nicer immediately, the following changes were made:

- Supress warnings and messages by default
- Set line spacing to one and a half (just looked it up in the internet)
- Do not show the whole table in the beginning but only the first lines;
- do not show the R code in this context since it is not meaningful;
- use `knitr::kable()` to print tables

- Do not show the code for preparing the plot, it is not necessary to understand the message
- Adjust `out.width` and `out.height` options in the plot chunk such that the plot is easier to read
- Show the code use to compute the correlation coefficient, but in a readable way; but summarize the output concisely, focusing on what is relevant
- Report the result of the Pearson correlation test in a more concise way

Of course, the last sentence above is true: to analyze this data in a meaningful way, we must invest a bit more thinking into the correct analysis method!