Visualize Data with ggplot2

THE UNIVERSITY of TULSA

AACSB ACCREDITED

DCU BUSINESS SCHOOL

DCU Dublin City University

dotLAB The Irish Institute of Digital Business

ggplot2, a major part of the Tidyverse

ggplot2 is the second version of the package ggplot (which no longer exist)

It stands for "Grammar of Graphics"

It is the most used package to plot data

```
install.packages("ggplot2")
```

```
library(ggplot2)
```

First steps

We need some data!

```
data_managers <- read.csv("https://bit.ly/2HvAuAO")
```

but these data need some wrangling

```
library(dplyr)

data_managers <- data_managers %>%
  rename(
    MCN = monthly.contract.negociated,
    JS_Q1 = job.satisfaction.Q1,
    JS_Q2 = job.satisfaction.Q2,
    JS_Q3 = job.satisfaction.Q3,
    JS_Q4_Rev = job.satisfaction.Q4.reverse,
    JS_Q5 = job.satisfaction.Q5
  ) %>%
  mutate(JS_Q4 = (10-0) - JS_Q4_Rev) %>%
  mutate(JS_score = cbind(JS_Q1,JS_Q2,JS_Q3,JS_Q4,JS_Q5) %>% rowMeans())
```

Layers of ggplot2

Describes all the non-data ink	Theme
Plotting space for the data	Coordinates
Statistical models & summaries	Statistics
Rows and columns of sub-plots	Facets
Shapes used to represent the data	Geometries
Scales onto which data is mapped	Aesthetics
The actual variables to be plotted	Data

Using ggplot2

First, we need to initiate the plot

```
ggplot(data = data_managers)
```

Then, we need to add some layers

- points, lines, columns, boxplots, ...
- axis characteristics
- plot specifications

Data layers (points, lines, columns, boxplots, ...)

All functions are starting with `geom_` in ggplot2

- points: `geom_point()`
- lines: `geom_line()`
- columns: `geom_col()`
- boxplots: `geom_boxplot()`

Arguments

- mapping = aes(x = Xcol, y = Ycol, color = Zcol, fill = Zcol, size = Zcol)
- size
- color
- ...

Examples with x = managers and y = MCN

```
ggplot(data = data_managers) +
  geom_point(mapping = aes(x = managers, y = MCN))

ggplot(data = data_managers) +
  geom_line(mapping = aes(x = managers, y = MCN))

ggplot(data = data_managers) +
  geom_col(mapping = aes(x = managers, y = MCN))

ggplot(data = data_managers) +
  geom_boxplot(mapping = aes(x = managers, y = MCN))
```

Examples with x = JS_score and y = MCN

```
ggplot(data = data_managers) +
  geom_point(mapping = aes(x = JS_score, y = MCN))

ggplot(data = data_managers) +
  geom_line(mapping = aes(x = JS_score, y = MCN))

ggplot(data = data_managers) +
  geom_col(mapping = aes(x = JS_score, y = MCN))

ggplot(data = data_managers) +
  geom_boxplot(mapping = aes(x = JS_score, y = MCN))
```

geom_smooth(), a special summary layer

Arguments

- method = "lm", "glm", "loess" or "gam"

```
ggplot(data = data_managers) +
  geom_point(mapping = aes(x = JS_score, y = MCN)) +
  geom_smooth(mapping = aes(x = JS_score, y = MCN), method = "lm")
```

Much better!

let's add some colors

```
ggplot(data = data_managers) +
  geom_point(mapping = aes(x = JS_score, y = MCN, color = managers)) +
  geom_smooth(mapping = aes(x = JS_score, y = MCN, color = managers),
              method = "lm",
              fullrange = TRUE)
```

Axis characteristics layers

- scale_x_continuous() / scale_x_discrete()
- scale_y_continuous() / scale_y_discrete()
- scale_color_continuous() / scale_color_discrete() / scale_color_manual()

```
ggplot(data = data_managers) +
  geom_point(mapping = aes(x = JS_score, y = MCN, color = managers)) +
  geom_smooth(mapping = aes(x = JS_score, y = MCN, color = managers),
              method = "lm",
              fullrange = TRUE) +
  scale_x_continuous(name = "Job Satisfaction Score") +
  scale_y_continuous(name = "Monthly Contract Negociated (avg.)", limits = c(0,NA)) +
  scale_color_manual(name = "Manager Experience", values = c("red","blue"))
```

Plot specifications layers

Facetting with facet_wrap() and facet_grid()

```
ggplot(data = data_managers) +
  geom_point(mapping = aes(x = JS_score, y = MCN, color = managers)) +
  geom_smooth(mapping = aes(x = JS_score, y = MCN, color = managers),
              method = "lm",
              fullrange = TRUE) +
  scale_x_continuous(name = "Job Satisfaction Score") +
  scale_y_continuous(name = "Monthly Contract Negociated (avg.)") +
  scale_color_manual(name = "Manager Experience", values = c("red","blue")) +
  facet_wrap(~ managers)
```

Plot specifications layers

Themes with theme_bw(), theme_minimal(), ...

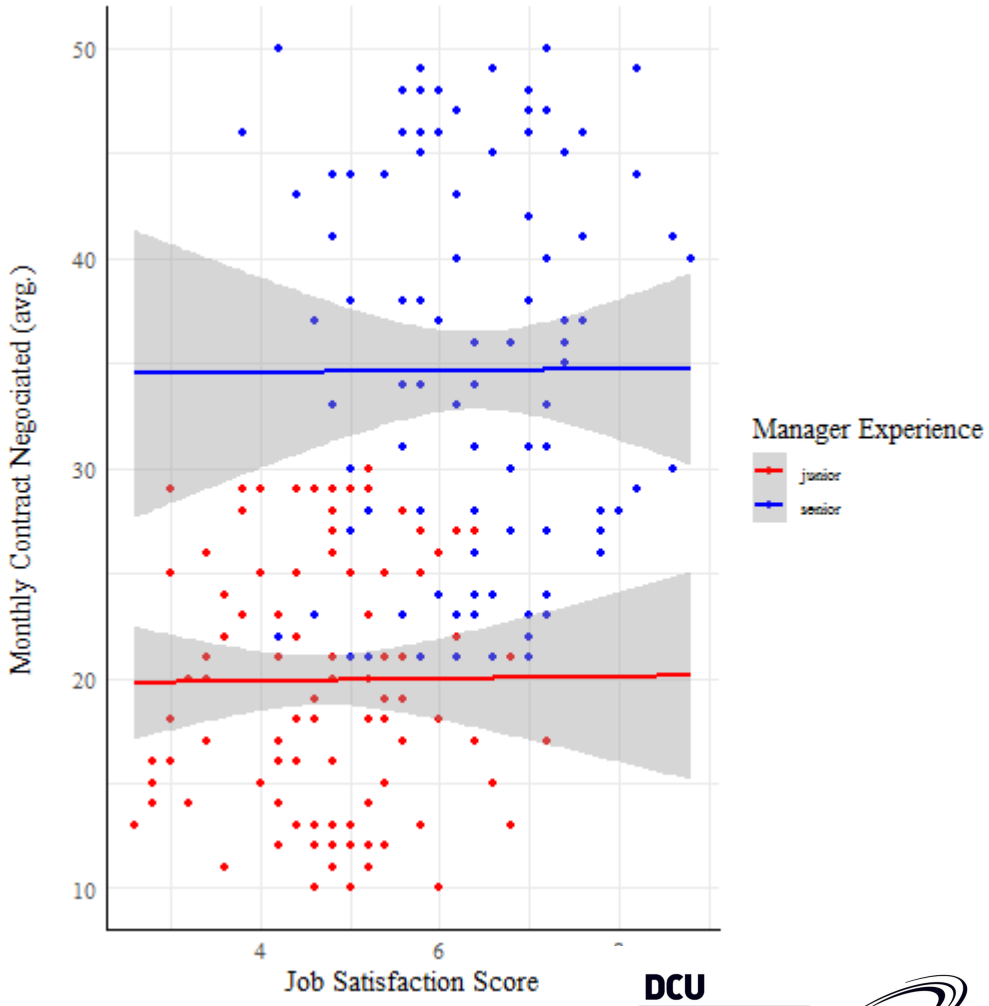
see all the basic themes here: https://www.r-graph-gallery.com/192-ggplot-themes/

```
ggplot(data = data_managers) +
  geom_point(mapping = aes(x = JS_score, y = MCN, color = managers)) +
  geom_smooth(mapping = aes(x = JS_score, y = MCN, color = managers),
              method = "lm",
              fullrange = TRUE) +
  scale_x_continuous(name = "Job Satisfaction Score") +
  scale_y_continuous(name = "Monthly Contract Negociated (avg.)") +
  scale_color_manual(name = "Manager Experience", values = c("red","blue")) +
  theme_minimal()
```

More specifications

```
ggplot(data = data_managers) +
  geom_point(mapping = aes(x = JS_score, y = MCN, color = managers)) +
  geom_smooth(mapping = aes(x = JS_score, y = MCN, color = managers),
              method = "lm",
              fullrange = TRUE) +
  scale_x_continuous(name = "Job Satisfaction Score") +
  scale_y_continuous(name = "Monthly Contract Negociated (avg.)") +
  scale_color_manual(name = "Manager Experience", values = c("red","blue")) +
  theme_minimal() +
  theme(
    text = element_text(size=14, family="serif"),
    axis.line.x = element_line(color="black", size = 0.1),
    axis.line.y = element_line(color="black", size = 0.1),
    axis.title.y = element_text(margin = margin(t = 0, r = 20, b = 0, l = 0)),
    legend.text = element_text(size=8),
    legend.position = "bottom")
```

Which gives...


Monthly Contract Negociated (avg.)
50
40
30
20
10
4
6
Job Satisfaction Score
Manager Experience
junior
senior

More possibilities with ggplot

Have a look at the ggplot Cheat Sheet

https://www.rstudio.com/wp-content/uploads/2015/03/ggplot2-cheatsheet.pdf

or google "ggplot cheat sheet"

Your are now a master of ggplot!