

# Speak visually!

## Data visualization in R using ggplot2

Pick a project, download the data, and choose a question. If you have a more interesting question than try to answer for your own question :)

Feel free to work on your own data.

Feel free to ask me!

### Useful webpages

Randomize nice colours: <https://coolers.co/app>

ggplot cheat sheet online: <https://www.rstudio.com/wp-content/uploads/2015/03/ggplot2-cheatsheet.pdf>

Quick help:

Scatterplot: <http://www.sthda.com/english/wiki/ggplot2-scatter-plots-quick-start-guide-r-software-and-data-visualization>

Histogram: <http://www.sthda.com/english/wiki/ggplot2-histogram-plot-quick-start-guide-r-software-and-data-visualization>

Bar chart: <http://www.sthda.com/english/wiki/ggplot2-barplots-quick-start-guide-r-software-and-data-visualization>

### Projects

#### Crime rate and weather in Boston and Los Angeles

Data: weather\_crimes\_Boston.csv and weather\_crimes\_LA.csv

The datasets contain the number of crimes per day from 2012 to 2017 in Boston and Los Angeles, also the type of the crime, and some weather measurement, such as the temperature, wind, snow, and rain.

Source: <https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/4UZ9D4>

Questions:

- Is it true that in the rainy weather nobody want to break into your house?
- Where should I move if I prefer good public safety?
- When do vandals like to “work”?
- When is it safer to be on the street, on weekdays or on weekend?
- On which day is it more likely that somebody will break into my house?
- Do burglars go on holidays?

#### Basketball players

Data: height\_basket\_vs\_football.csv

Data contain selected players' height in feet. 1 feet = 30.48 cm

Source: NBA Encyclopedias,

[https://college.cengage.com/mathematics/brase/understandable\\_statistics/7e/students/datasets/tvds/frames/frame.html](https://college.cengage.com/mathematics/brase/understandable_statistics/7e/students/datasets/tvds/frames/frame.html)

Questions:

- Show in a boxplot that who is taller: an average basketball or football player?
- Are you taller than an average basketball player?

### Research and innovation expenses in Hungary

Data: ksh\_publication.csv and ksh\_research

These datasets contain the amount of money (in million forints) was given for research purposes in Hungary in the last 20 years and the published papers and books.

Source: KSH, [http://www.ksh.hu/kutatas\\_fejlesztés\\_innováció](http://www.ksh.hu/kutatas_fejlesztés_innováció)

Questions:

- How much money did Hungary spend on research and innovation?
- How is the amount divided between universities, research institutes and companies?
- Is it beneficial to spend on researching according to the publications?
- Hungary joined to the EU in 2003. Is it visible on the money spent on researching and innovation?

### Data around the world

Data: this is a package. You can download and read with this code: (you also find that script in the speak\_visually.R file)

```
if (!("devtools" %in% installed.packages())){
  install.packages(devtools)
}
devtools::install_github("tadaadata/loldata")
library(loldata)

str(worldrankings)
worldrankings$happiness
worldrankings$smartphone_adoption
```

More information about this dataset and sources: <https://github.com/tadaadata/loldata>

The dataset contains fun information about countries, such as the rate of happiness, smartphone addiction, gender equality and more.

Questions:

- Is there a correlation between happiness and smartphone addiction?
- Are people who have better education happier?
- Are people more satisfied in countries where gender equality is more balanced?
- Which is the most peaceful country according to the dataset?
- Where is highest the life expectancy?

### Late night tweet

Data: late\_night\_tweet.csv

The dataset contains basketball players' information about whether they tweeted at late night before a game and their performance in the game. To understand the scores in basketball check this out:

[https://en.wikipedia.org/wiki/Basketball\\_statistics](https://en.wikipedia.org/wiki/Basketball_statistics)

Source: <https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/JMPGZX>

Questions:

- Who is the best basketball players according to the dataset?
- Should the trainer forbid late-night tweets?

## Climate change

Data:

- temperature.csv, Data are included from the GISS Surface Temperature (GISTEMP) analysis and the global component of Climate at a Glance (GCAG).  
source: <https://github.com/datasets/global-temp>
- climate.csv, climate\_northern.csv, climate\_sou.csv, Source: <https://data.giss.nasa.gov/gistemp/>  
global average temperature, the average temperature in the northern and in the southern hemisphere.

Questions:

- How the average winter temperature is changing?
- Is the effect of the industrial revolution visible on the climate according to the data?
- Are the temperature changes in the northern and the southern hemisphere similar?

## More

data visualization:

- <https://informationisbeautiful.net>
- <https://venngage.com/blog/how-to-choose-the-best-charts-for-your-infographic/>
- <https://www.youtube.com/watch?v=5Zg-C8AAIGg> – David McCandless' ted talk about data visualization
- [https://eazybi.com/blog/data\\_visualization\\_and\\_chart\\_types/](https://eazybi.com/blog/data_visualization_and_chart_types/)
- <http://junkcharts.typepad.com/>
- [www.storytellingwithdata.com/](http://www.storytellingwithdata.com/)
- <https://www.tableau.com/learn/articles/best-beautiful-data-visualization-examples>
- <https://datavizcatalogue.com/>
- <https://www.youtube.com/watch?v=8EMW7io4rSI> – story telling with data, by Cole Nussbaumer Knaflic
- <https://www.youtube.com/channel/UCjhGIILWNloXJdR2NTCBMIA> - Cole Nussbaumer Knaflic's youtube channel
- <https://www.youtube.com/watch?v=hVimVzgtD6w> he best stats you've ever seen. Ted talk of Hans Rosling

misleading graphs:

- <https://venngage.com/blog/misleading-graphs/>
- <https://www.mediamatters.org/research/2012/10/01/a-history-of-dishonest-fox-charts/190225>

play around:

- <http://www.tylervigen.com/spurious-correlations>