```
library(dplyr)

rladies_global %>%
  filter(city == 'Cape Town')
```

Sharing experiences: Data visualization & R-Ladies Tbilisi

Liili Abuladze, 4 November 2017

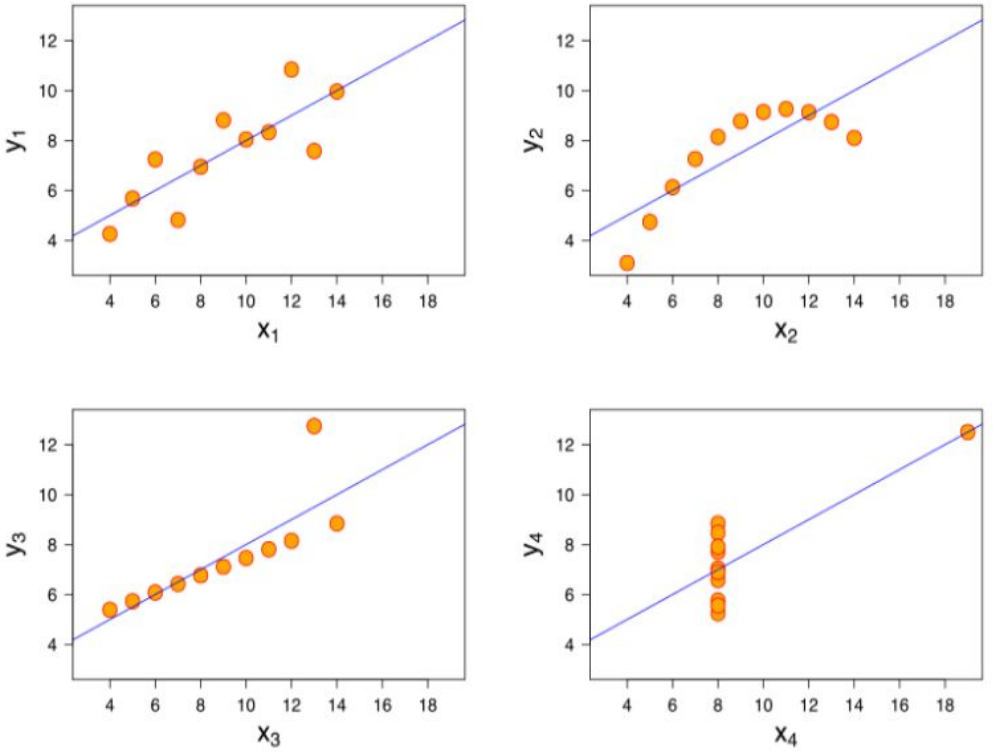
Why plot data?

- Visual presentation of numbers and data points
- Get our message across by simplifying complex data
  - Let the data focus and get to the point
- Explore and understand data before analyzing it
- Check inconsistencies
- Find outliers

Anscombe's quartet

4 datasets with same descriptive stats:

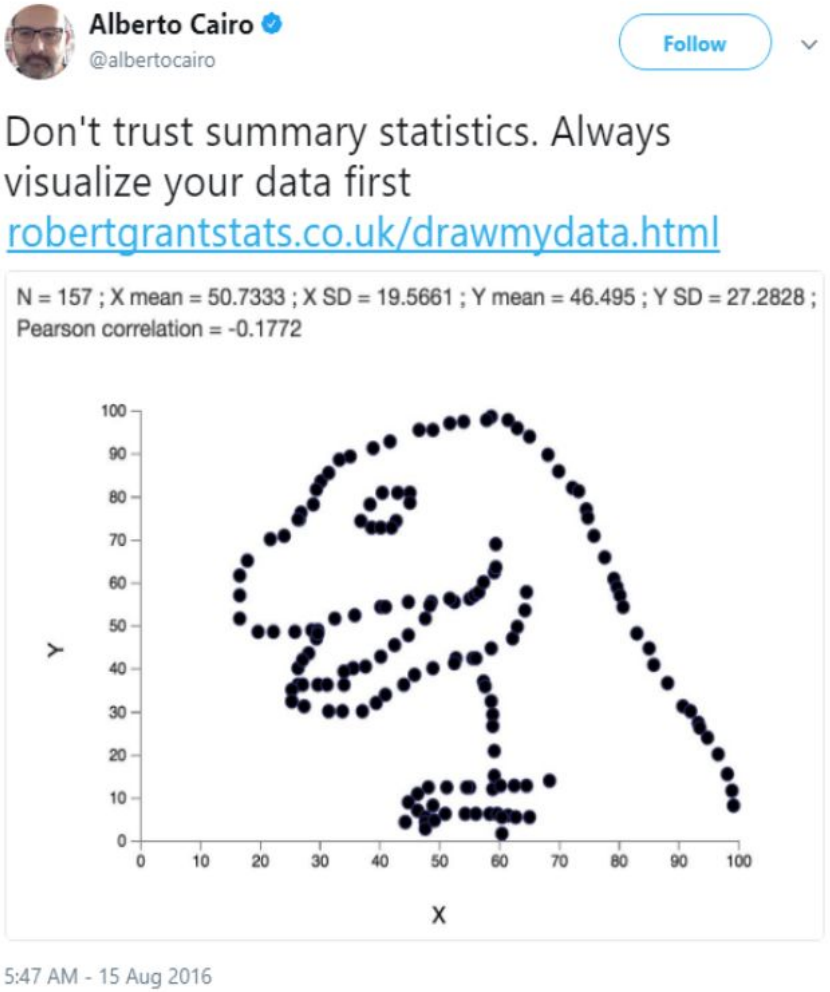
- Top panel: high correlation even if nonlinear relationship
- Below: summary statistics are sensitive to outliers



Source: [Wikipedia]

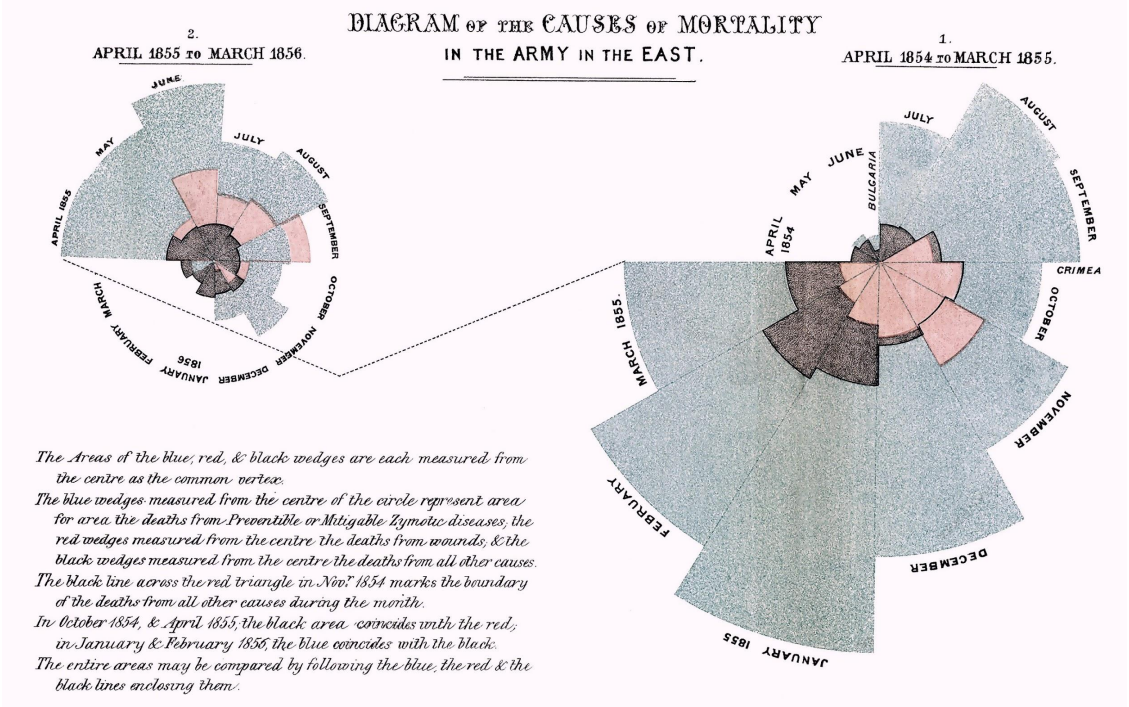
Don’t trust summary statistics…

Source: Alberto Cairo’s tweet



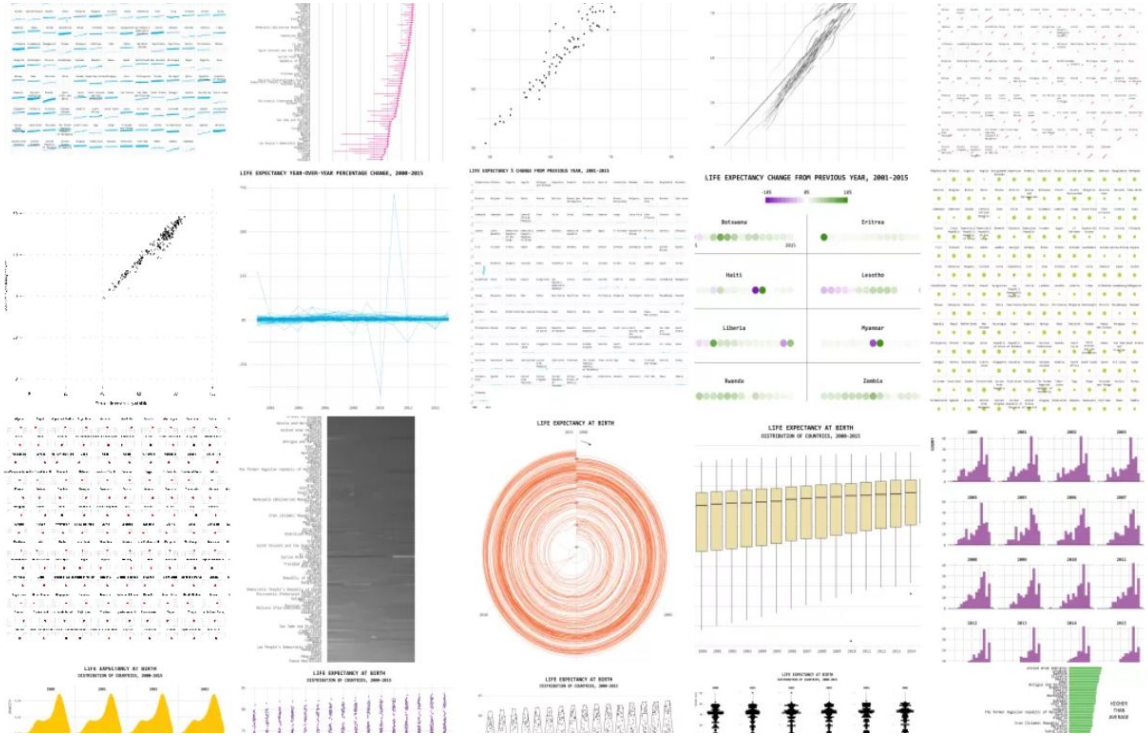
See the data clearly, and act on it

- “Polar area diagrams” of causes of death in the Crimean war, 1858, by Florence Nightingale
- Indicated that most soldiers died of preventable causes (sanitation), in blue



Source: Oscar technology website

Get your message across



Source: [Flowingdata - One Dataset, Visualized 25 Ways]

Edward Tufte's principles of graphical excellence

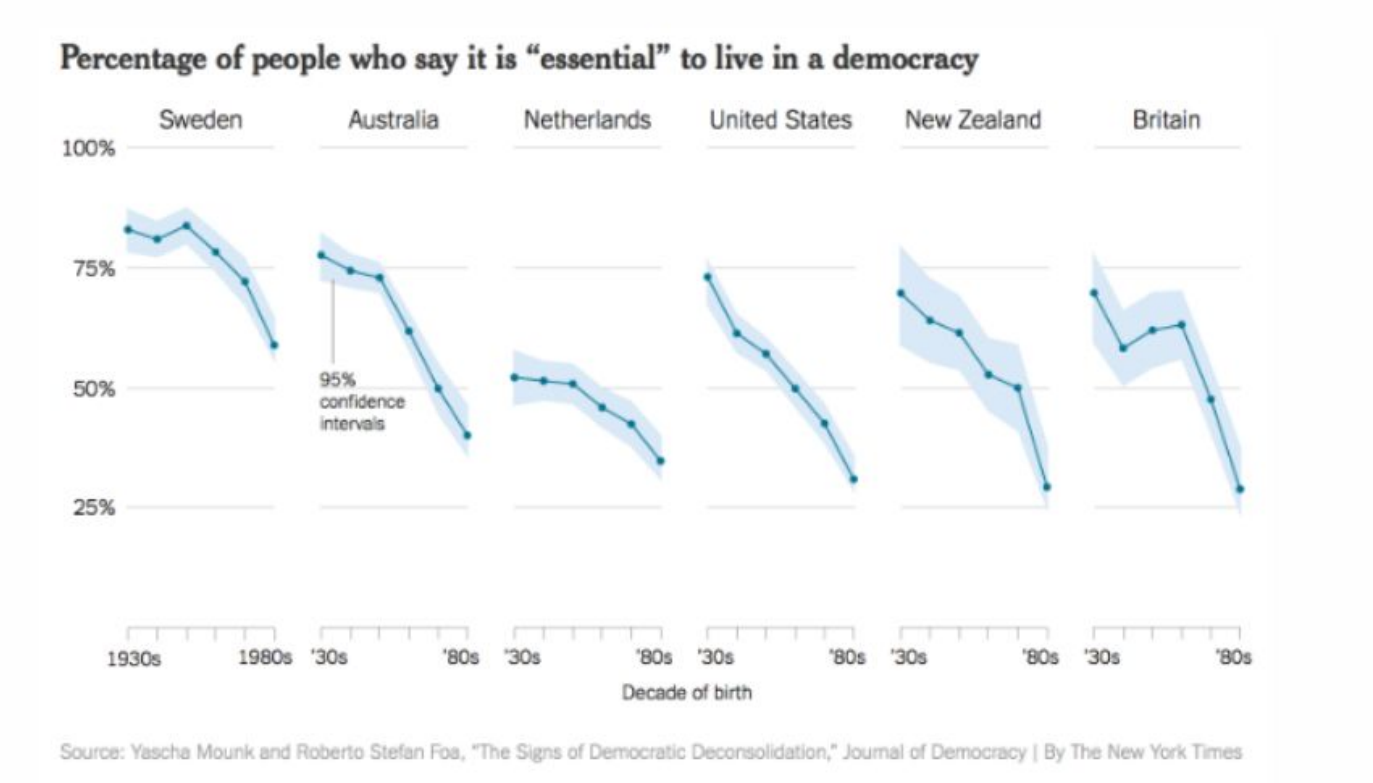
- well-designed presentation of interesting data - a matter of substance, of statistics, and of design
- complex ideas communicated with clarity, precision and efficiency
- which gives to the viewer the greatest number of ideas in the shortest time with the least ink in the smallest space
- [It] is nearly always multivariate
- requires telling the truth about the data

Source: Tufte 1983, from Kieran Healy's "Data Visualization for Social Science"

Edward Tufte's principles of graphical excellence

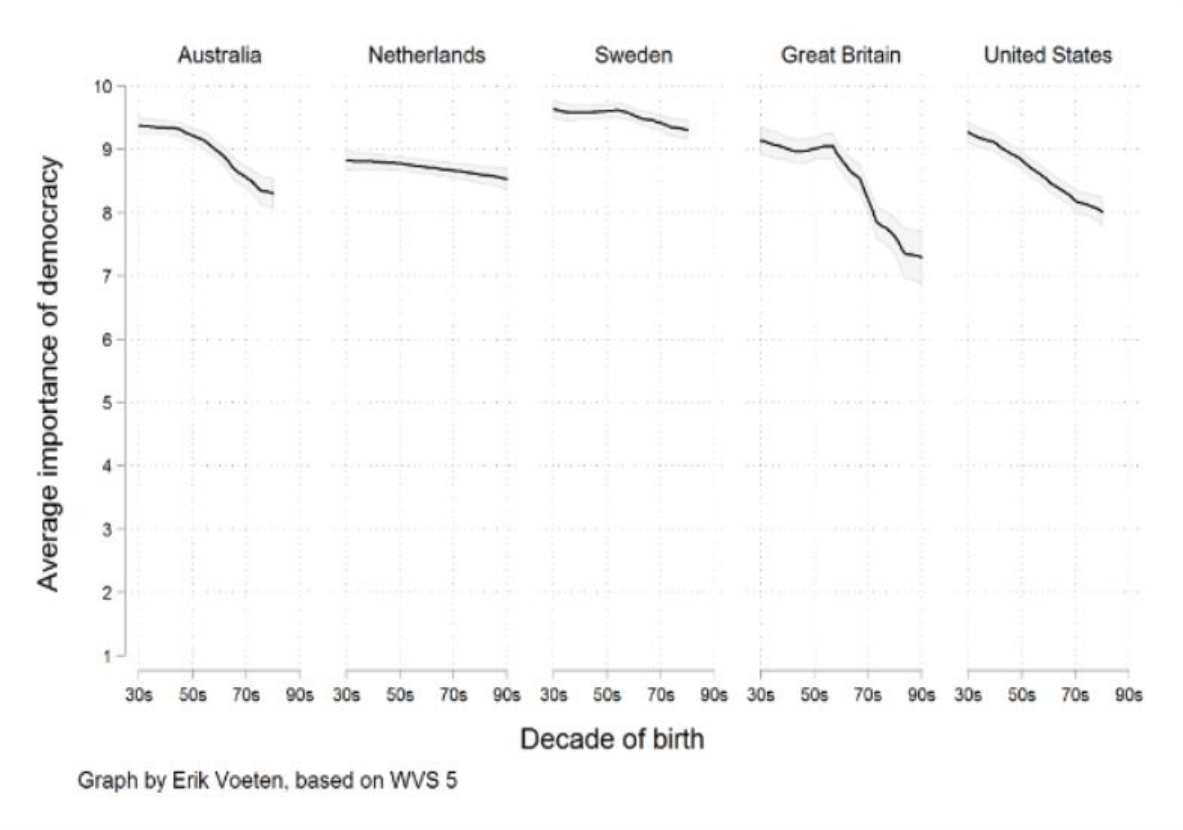
- Clear, detailed, thorough labelling with proper scales
- Graphic effect size should be proportional to the actual effect present in the data ("lie factor")
- Maximize data-to-ink ratio
- Avoid chart junk

A crisis of faith in democracy? (New York Times)



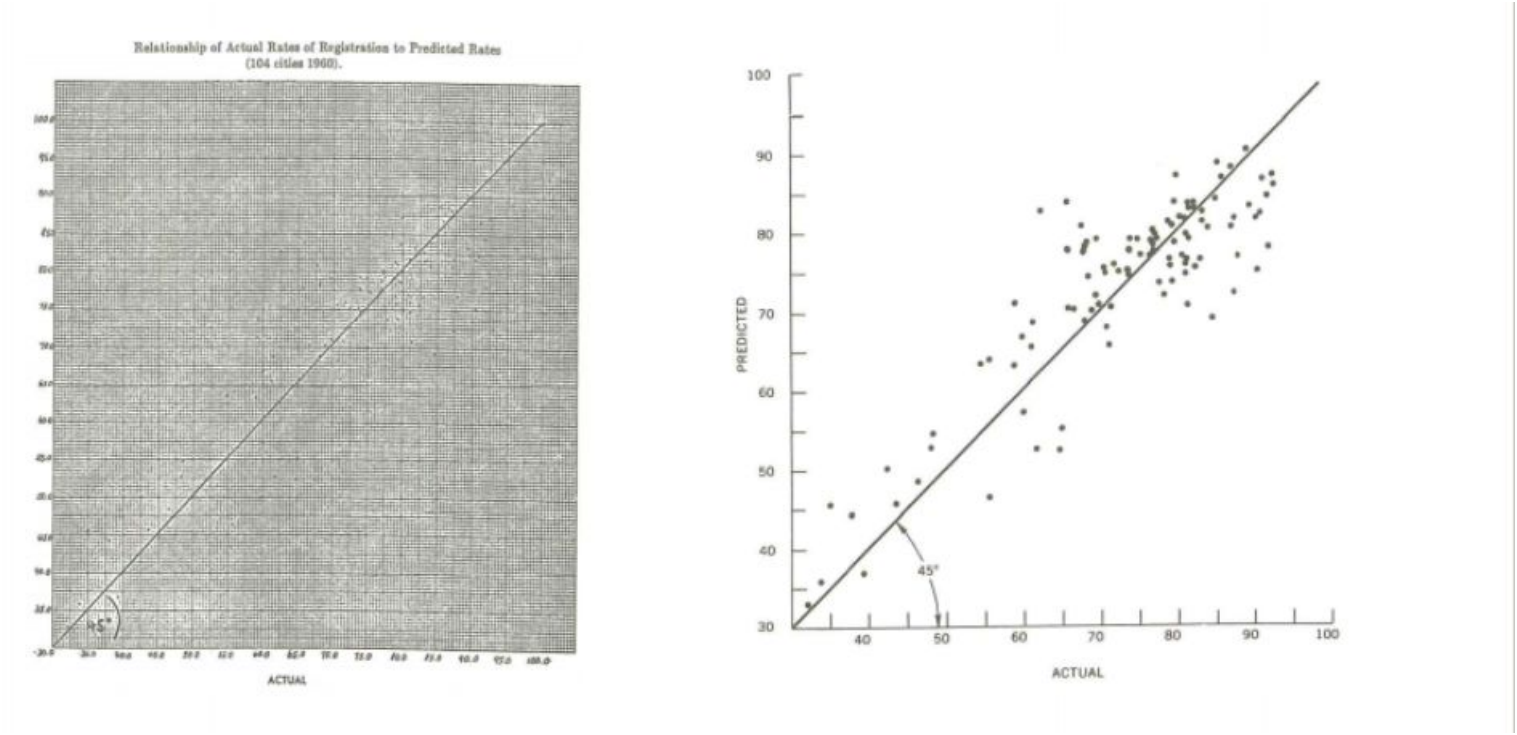
Source: http://socviz.co/lookatdata.html

An overblown crisis



Source: http://socviz.co/lookatdata.html

Data-to-ink ratio and chartjunk



Source: http://classes.engr.oregonstate.edu/eecs/spring2015/cs419-001/Slides/tufteDesign.pdf

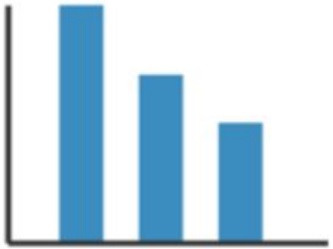
Not too minimal - gridlines and labels help to read data!

Perception

Figure 2.14: Mach bands. On the left side, five grey bars are ordered from dark to light, with gaps between them. On the right side, the bars have no gap between them. The brightness or luminance of the corresponding bars is the same. However, when the bars touch, the dark areas seem darker and the light areas lighter.

Source: http://socviz.co/lookatdata.html

Marks and attributes



Length

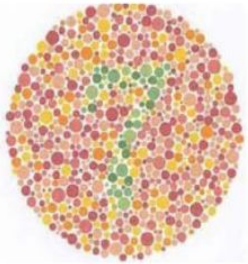
Position

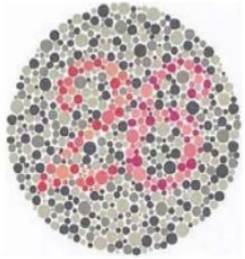
Color

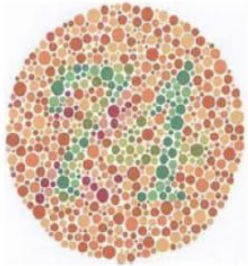
Size

Source:
http://saravanan-thirumuruganathan.github.io/cse5334Spring2015/slides/03_PrinciplesOfViz/03_PrinciplesOfViz_final.pdf

Color blindness







8% of males, 1% of females

Most common is red-green weakness /blindness

Source: http://saravanan-thirumuruganathan.github.io/cse5334Spring2015/slides/03_PrinciplesOfViz/03_PrinciplesOfViz_final.pdf

Color blindness

These are color-blind-friendly palettes, one with gray, and one with black.

To use with ggplot2, it is possible to store the palette in a variable, then use it later.

```
# The palette with grey:
cbPalette <- c("#999999", "#E69F00", "#56B4E9", "#009E73", "#F0E442", "#0072B2", "#D55E00", "#CC79A7")

# The palette with black:
cbbPalette <- c("#000000", "#E69F00", "#56B4E9", "#009E73", "#F0E442", "#0072B2", "#D55E00", "#CC79A7")

# To use for fills, add
  scale_fill_manual(values=cbPalette)

# To use for line and point colors, add
  scale_colour_manual(values=cbPalette)
```

Source: R cookbook

Resources

- Cynthia Brewer http://colorbrewer2.org/
- WTF visualizations http://viz.wtf/
- Kieran Healy's Data Visualization for Social Science http://socviz.co/
- Hadley Wickham's A layered grammar of graphics http://vita.had.co.nz/papers/layered-grammar.html
- Google group forum for ggplot2

Using ggplot2

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

```
library(ggplot)
```

- Data must be a dataframe!

```
library(gapminder)
```

```
is.data.frame(gapminder)
```

- Plot

```
ggplot(gapminder)
```

Adding layers - adding with a `+`

```
ggplot(data = gapminder) +
  geom_line(aes(x = year, y = lifeExp))
```

- Geometric objects

`geom_point()`

`geom_line()`

`geom_smooth()`

`geom_text()`

Aesthetics for shape, color, size, position

Mapping visual properties, such as `x`, `y`, `colour`, `size`, `shape`, `alpha`, `fill`, `radius`, `linetype`, `group`

- Do it inside the `(aes)`

```
ggplot(data = gapminder) +
  geom_line(aes(x = year, y = lifeExp, group = country))
```

Other elements

- Scale
- Grid
- Theme
- Stat
- Facet
- Coordinates

Tbilisi, Georgia

- Georgia made by characters

  https://www.youtube.com/watch?v=RhtEftOx-Gk&feature=youtu.be

- Population of Tbilisi 1,118, 035

R-Ladies Tbilisi

- ~ 30 meetups since December 2016
- A member survey:

https://goo.gl/forms/6lUzoKMNDOeJoGyo1

- September-October - beginners' workshop weekends:
  - Downloading & installing, basic operations (Liili Abuladze)
  - Data objects & structures & exploration (Liili Abuladze)
  - Data visualisation using ggplot2 (David Sichinava)
  - Basic statistics (Nino Melitauri)
  - Practical sessions (Tamuna Margievi, Nino, Liili, David)

Topics covered

- Reading & creating data
- Data wrangling, cleaning, manipulation
- Statistical analysis
- Spatial analysis, mapping
- Data visualisation, plotting
- Web scraping
- Machine learning

R-Ladies Tbilisi - next meetups

Meetups **@GeoLab**

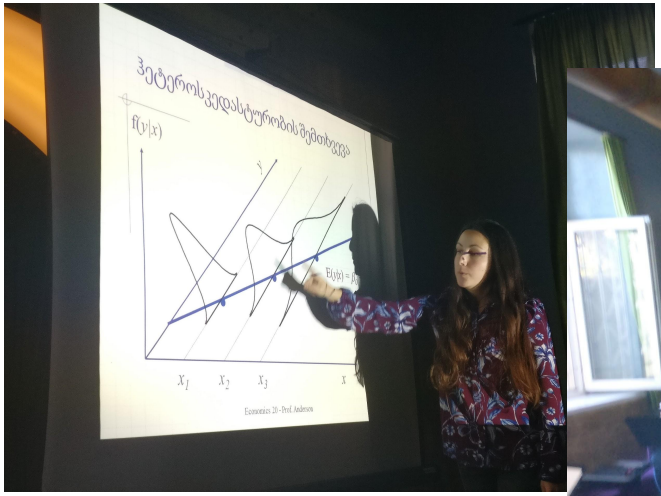
7 PM

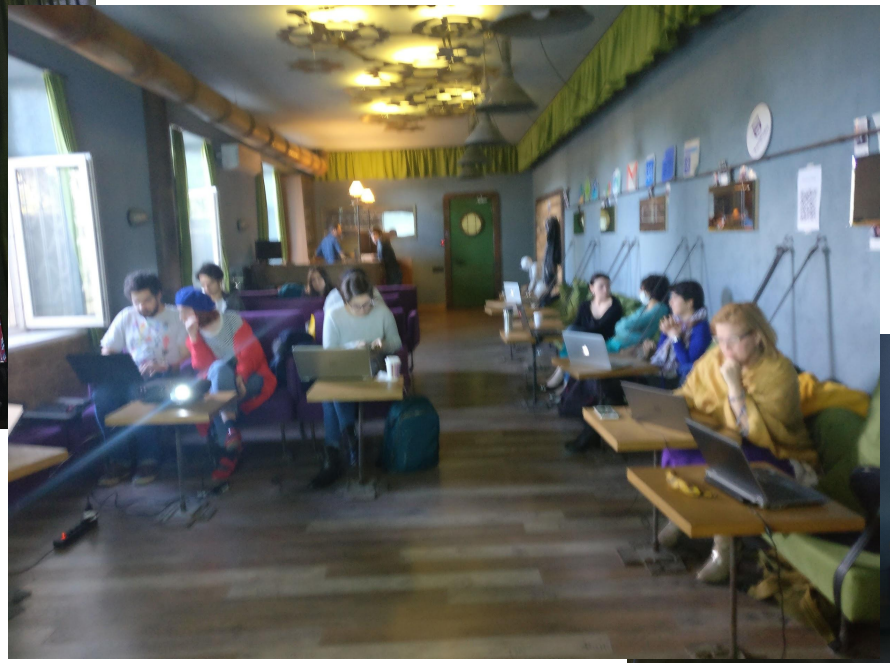
November 28 - Developing R packages

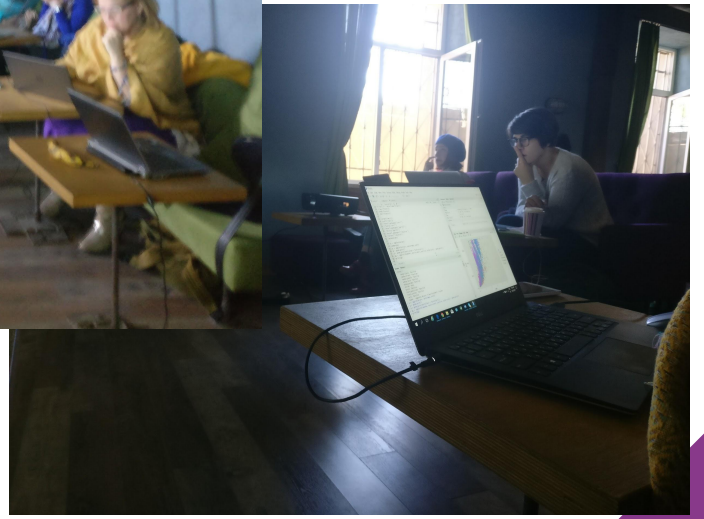
November-December - statistics, data visualisation, web scraping (?), etc.

Open for everyone, beginners welcome to drop in anytime

R-Ladies Tbilisi







Find us...

Events
https://www.meetup.com/rladies-tbilisi/

Materials !!
https://github.com/rladies/meetup-presentations_tbilisi

Social Media
Twitter @rladiestbilisi
Facebook @rladiestbilisi/

tbilisi@rladies.org