Introduction to Biostatistics

Leonardo Collado-Torres

December 26, 2016

What is Biostatistics?

Math?

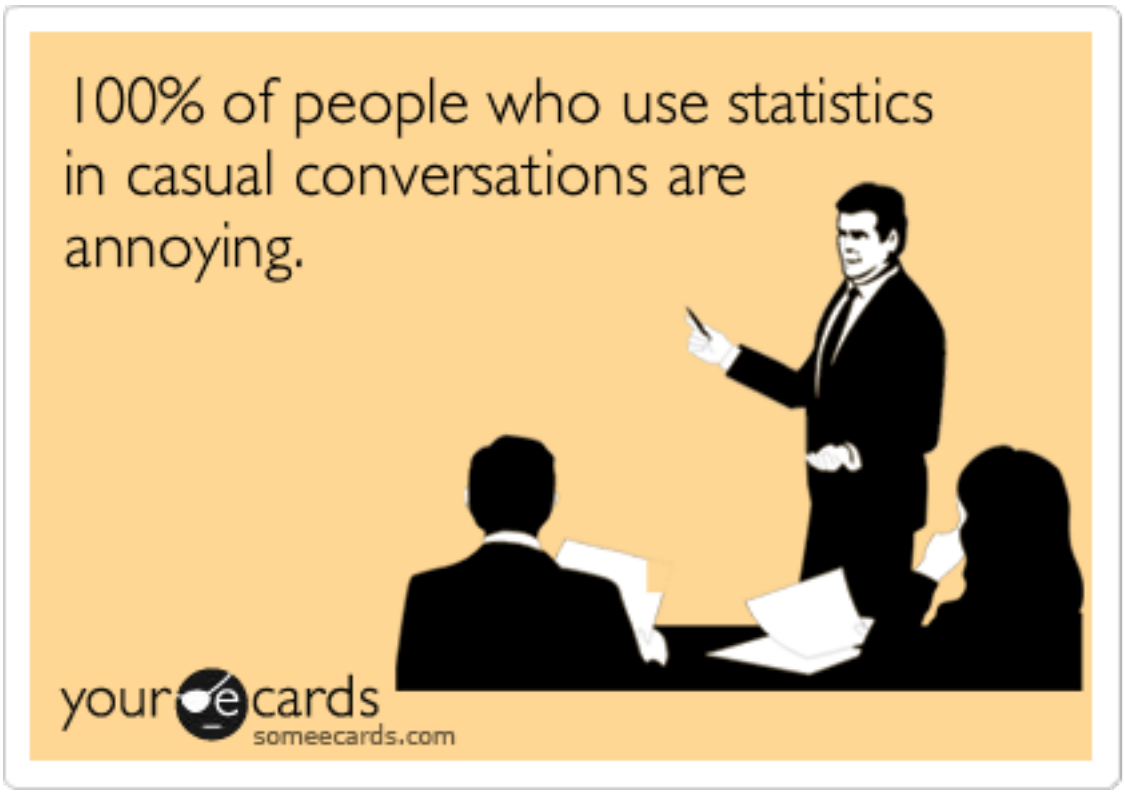
$$t = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{S_{\overline{X}_1 - \overline{X}_2}} = \frac{\overline{X}_1 - \overline{X}_2}{S_{\overline{X}_1 - \overline{X}_2}}$$

$$S_{\overline{X}_1 - \overline{X}_2} = \sqrt{\frac{(N_1 - 1)s_1^2 + (N_2 - 1)s_2^2}{N_1 + N_2 - 2}\left[\frac{1}{N_1} + \frac{1}{N_2}\right]}$$

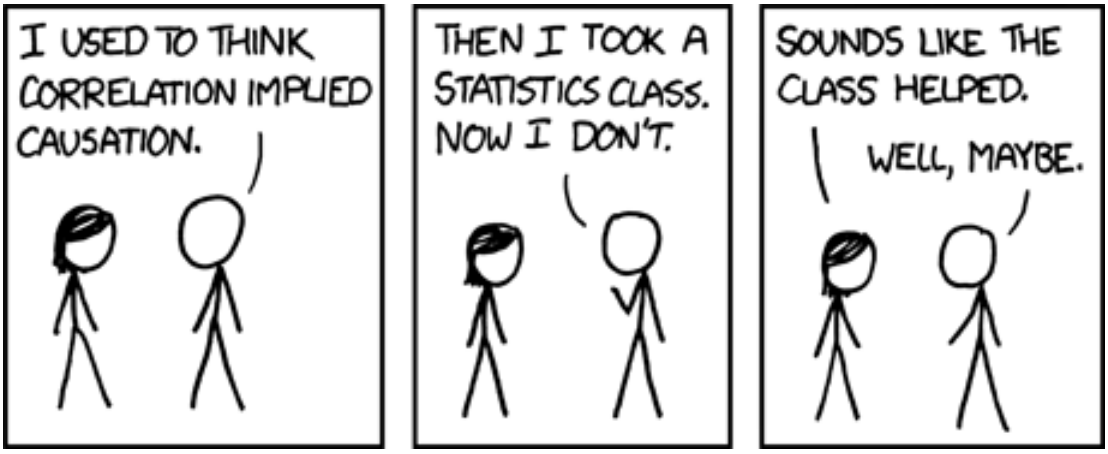
Source

Graphs?

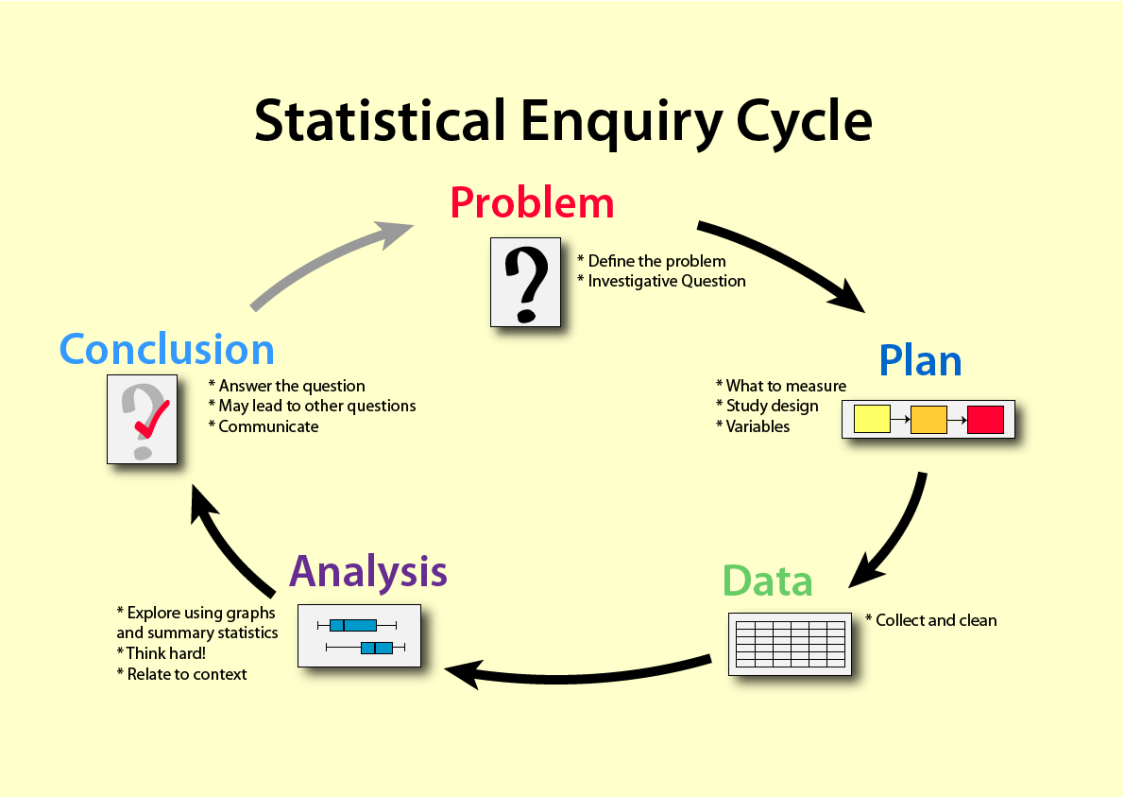
Source



Source



Source

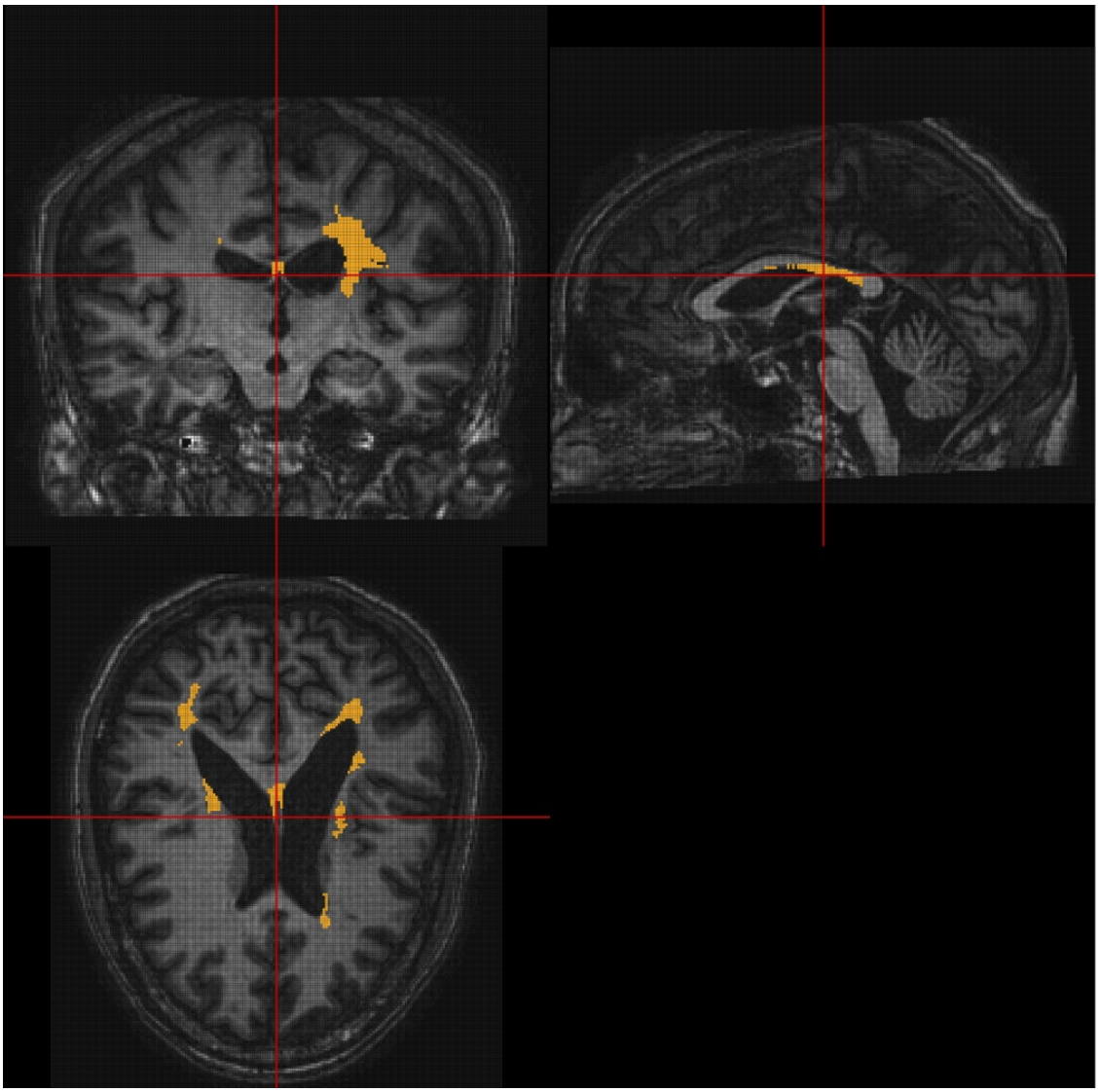


Source

Biostatistics

- Theoretical mathematics: building blocks for other researchers
- Applied math: use math to solve a problem like building an estimator
- Biostatistics; using stats with biological data

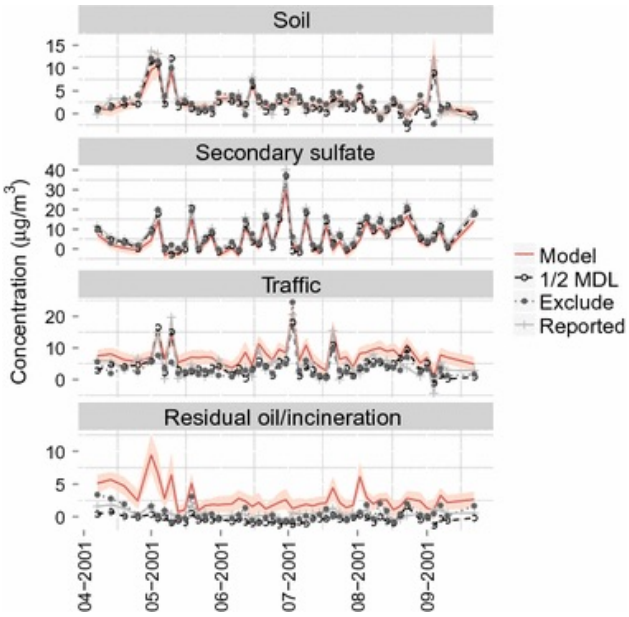
Brain imaging



Check out this interactive visualization!

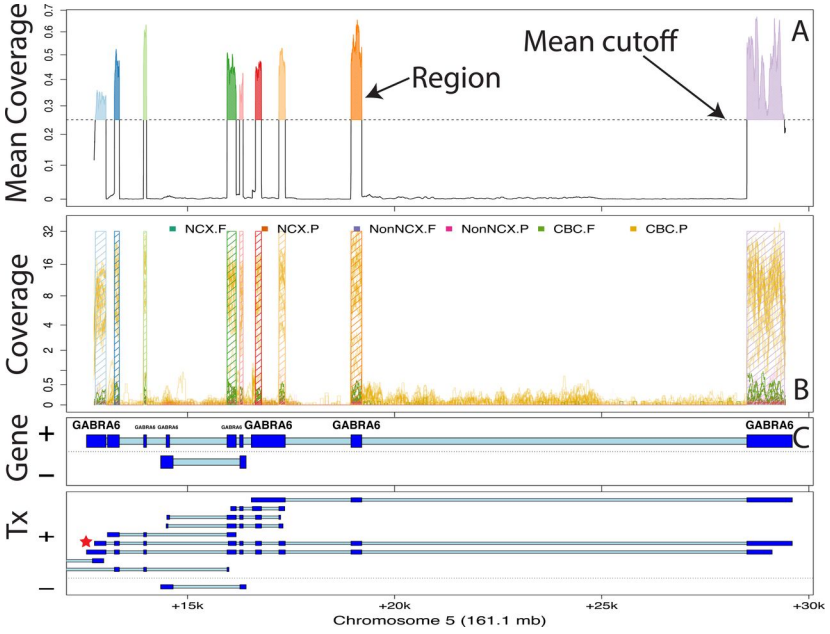
Source

Environmental Statistics



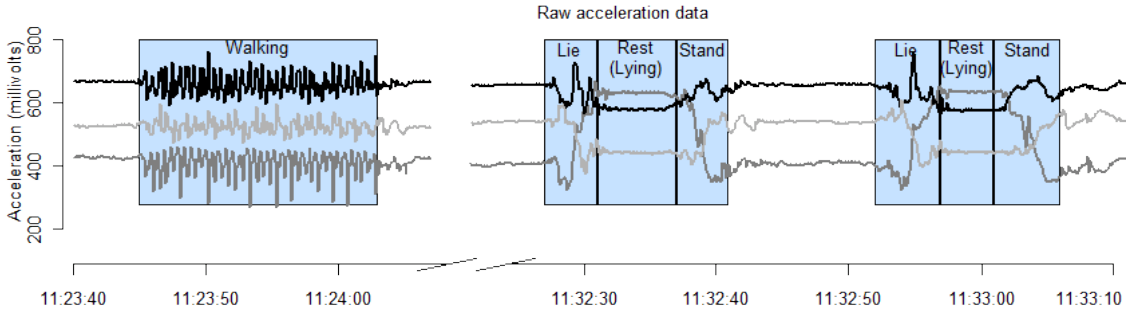
Source

Genomics



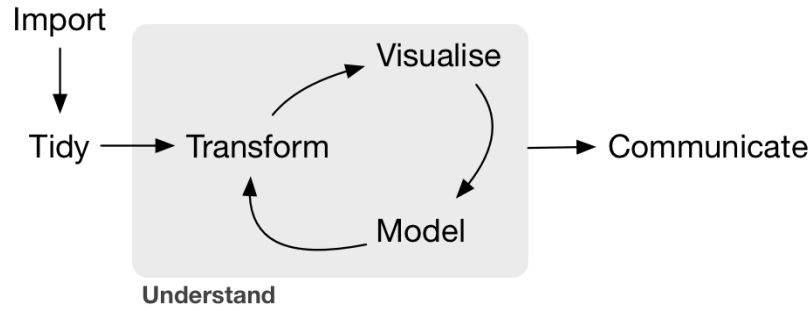
Source

Activity trackers



Source

Data Science overview



Source

Cleaning data

Source

One type of clean data: tidy data

from Wickham's *Tidy Data*

In **tidy data**:

1. Each variable forms a column.
2. Each observation forms a row.
3. Each type of observational unit forms a table.

messy

	treatmenta	treatmentb
John Smith	—	2
Jane Doe	16	11
Mary Johnson	3	1

	John Smith	Jane Doe	Mary Johnson
treatmenta	—	16	3
treatmentb	2	11	1

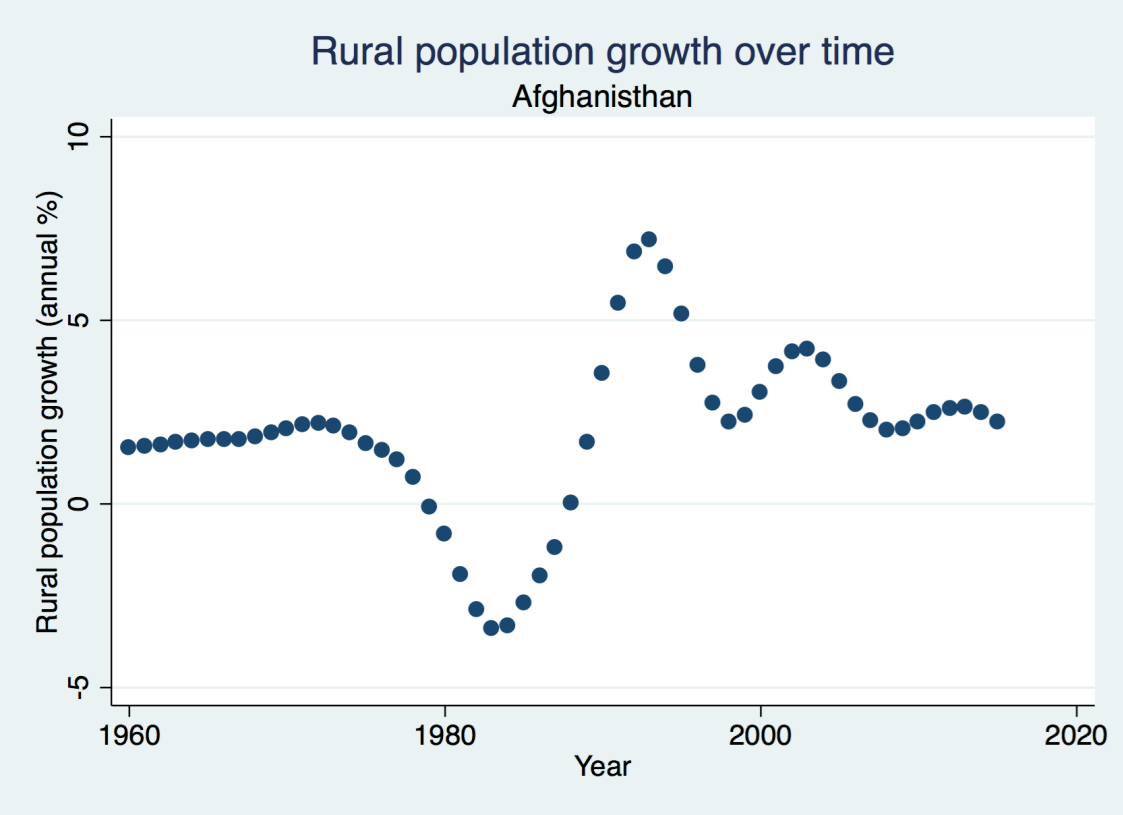
tidy

name	trt	result
John Smith	a	—
Jane Doe	a	16
Mary Johnson	a	3
John Smith	b	2
Jane Doe	b	11
Mary Johnson	b	1

Source

You'll learn Stata

```
use afg_worldbank_2016.dta
twoway (scatter SPRURTOTLZG Year), /*
    */ title(Rural population growth over time) /*
    */ subtitle(Afghanisthan)
```



Modern biostatistician

- Participates in the experimental design, that is the *question* the team wants to answer
- Is a full participant in the research
- Learns how to *clean* data
- Models the data given the question of interest
- Contributes in interpreting the data and suggesting the next steps

Communicating results

- Understand the main concepts so you can explain them: means understanding why we chose a particular method
- Proper graphics
- Interpretation of the results

Dr. McGready

Lets take a look at how John McGready introduces biostatistics

- Open lecture 1 of the course *Statistical Reasoning I*

Exercise

Divide in two teams and come up with a 2 minute overview of Biostatistics. Imagine that you will record a 2 minute video to get students excited about your Biostatistics course.

Jeff Leek Data Analysis intro

Source

Slides license