

# Colecionar Analizar Visualizar

Emily Dich + Raul Boquin

# Where do you get data? Donde puedes conseguir data?

- 1. Data Libraries Librarias de data
- 2. Web Scraping
- 3. Design your own experiment! disena su propio experimento

# 1) Data Librarias

http://data.worldbank.org/country/colombia

#### 2) Actividad - Web Scraping

#### **Installacion**

#### Install python

- a. <a href="https://www.python.org/downloads/">https://www.python.org/downloads/</a>
- b. 2.7
- c. click download link

#### 2. Install pip

- a. <a href="https://pip.pypa.io/en/stable/installing/">https://pip.pypa.io/en/stable/installing/</a>
- b. download get-pip.py

#### 3. Install jupyter

a. pip3 install jupyter

#### 2) Actividad - Web Scraping

- 1. pip install bs4
- 2. pip install requests
- 3. pip install pandas
- 4. pip install numpy
- 5. pip install matplotlib

### 3) Disena un experimento!

Survey monkey

Amazon mturk

<u>https://moves-app.com/</u> (apps that willing participants can install to track themselves)

Questionnaires in person on the street

A/B Testing

Organize a data collection team

## 3) Disena un experimento!

- 1) Data management plan
  - a) How will you keep it safe?
  - b) Will you be sharing?
  - c) Organization of the data? Tables? Etc
  - d) Collection of data?
  - e) Who? What? Why?
- 2) Obtain Human subject Approval
  - a) This is going into IRB approval
- 3) Design data collection instrument
- 4) Pilot the instrument
- 5) Implement

# You have data...now what? Ahora tienes data...y ahora que?

- 1) Analysis
- 2) Visualizacion

#### a. Quantitative

i.

#### a. Quantitative

- i. The frequency (rate, duration) of specific behaviors or conditions
- ii. Test scores (e.g., scores/levels of knowledge, skill, etc.)
- iii. Survey results (e.g., reported behavior, or outcomes to environmental conditions; ratings of satisfaction, stress, etc.)
- iv. Numbers or percentages of people with certain characteristics in a population (diagnosed with diabetes, unemployed, Spanish-speaking, under age 14, grade of school completed, etc.)

#### a. Quantitative

- i. La frecuencia de un condicion particular
- ii. Markas de examen
- iii. Resultados de una encuesta
- iv. Numeros de percentajes de personas con características

٧.

#### a. Qualitative

i.

#### a. Qualitative

- i. qualitative information tends to be "soft," meaning it can't always be reduced to something definite. That is in some ways a weakness, but it's also a strength. A number may tell you how well a student did on a test; the look on her face after seeing her grade, however, may tell you even more about the effect of that result on her.
- ii. Qualitative data can sometimes be changed into numbers, usually by counting the number of times specific things occur in the course of observations or interviews, or by assigning numbers or ratings to dimensions subjective

#### 1) Analysis - Actividad

#### a. Analyze your Facebook Information

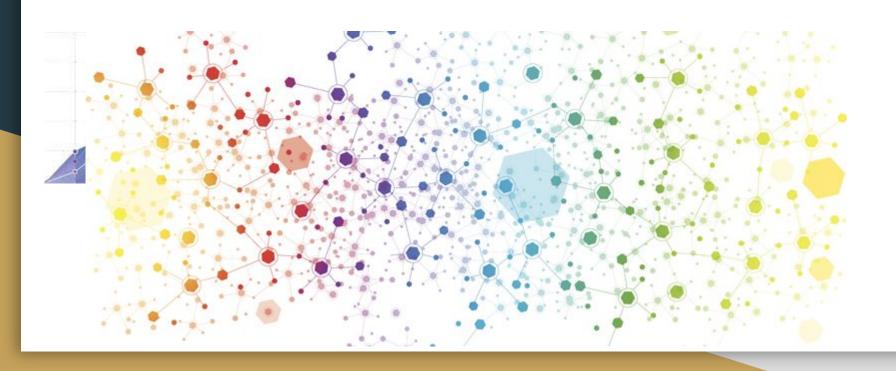
- i. <a href="http://www.wolframalpha.com/input/?i=facebook+report">http://www.wolframalpha.com/input/?i=facebook+report</a>
- b. Go to this site Vaya a este link
- c. Make a wolfram user account and connect your facebook to it Haga una cuenta de Wolfram y conecta su Facebook
- d. Look at the data analysis that has taken place mira lo que pasa!
- e. Then, discuss with your group the importance of using it or why it exists.
  - i. discuta con un companero la importancia de este arramiento.
- f. How did Wolfram organize and analyze your data?
  - i. Como Wolfram organiza y analiza su data de Facebook?
- g. What are the implications from its analysis?
  - Que son las implicaciones de este analysis
- h. How is it represented?
  - i. Come decide Wolfram visualizar y representat este data?

#### 1) Analysis - Actividad

a. <a href="http://tylervigen.com/spurious-correlations">http://tylervigen.com/spurious-correlations</a>

b.

#### Visualizacion - contando una estoria



#### The chart selector — some basic chart suggestions Variable Width Table With Column Chart Circular Area Column Chart Line Chart **Bar Chart** Line Chart Column Chart Embedded Charts Chart Many Items Few Items Single/Few Cyclical Non-Cyclical Many Two Variables Many Few Data Categories Categories Data Categories per Item Categories One Variable Many Few per Item Periods Periods Column Among Items Histogram Over Time Few Data Points Scatterplot Comparison Single Two Variable Variables Line Histogram Many What would you like Data Relationship Distribution Points to show? **Bubble Chart** Three Variables Composition Scatterplot Two Variables Changing Over Time 3D Area Chart Static Three Variables -Many Periods Few Periods Only Relative Relative & Absolute Only Relative Relative & Absolute Simple Share Accumulation or Components Differences Matter Differences Matter of Total Subtraction to Total of Components Differences Matter Differences Matter Stacked 100% Stacked Column Stacked 100% Stacked Area Stacked 100% Pie Chart Waterfall Chart Column Chart Chart Area Chart Column Chart with Chart Subcomponents @ Randal S. Olson Inspired by work by Dr. A. Abela

Emily - ydich@college.harvard.edu

Raul - boquinr@mit.edu