DATA ANALYSIS

Bryn Mawr Digital Scholarship

Graduate Community of Learning

March 27, 2018

RAWGraphs

WHAT IS DATA ANALYSIS?

THE FOUR PHASES OF DATA ANALYSIS

PHASE 1: REFINE YOUR QUESTION

- Are air pollution levels higher on the east coast than on the west coast?

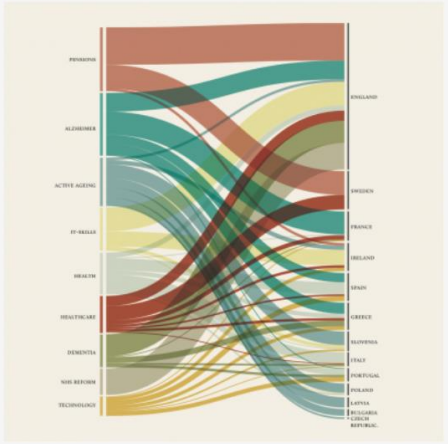
- Are hourly ozone levels on average higher in New York City than they are in Los Angeles?

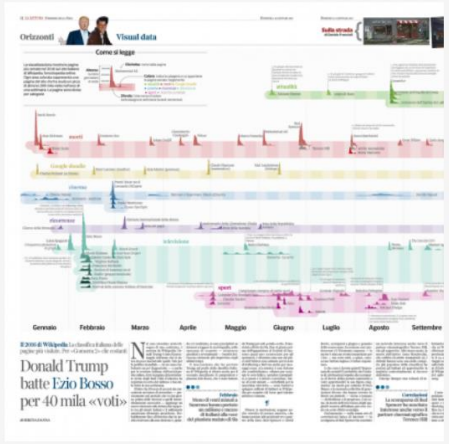
PHASE 2: EXPLORE THE DATA

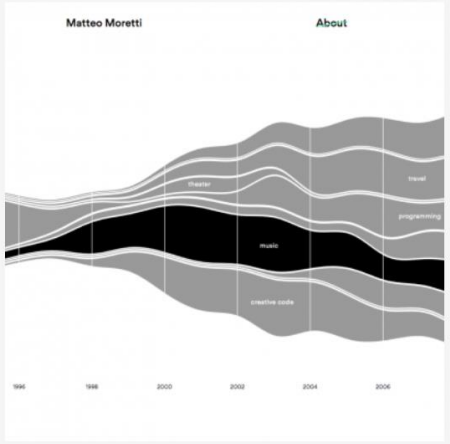
Goals of Exploratory Data Analysis:

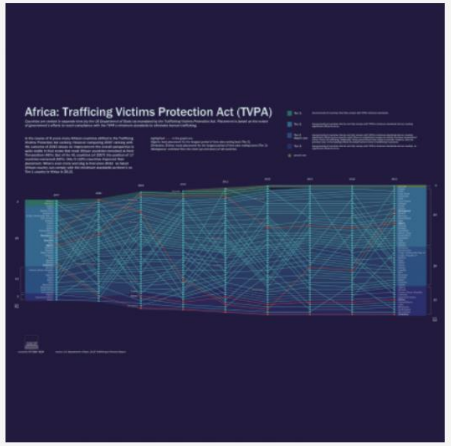
- Summarize and understand the structure of a dataset
- Identify any assumptions you have about the data
- Formulate some hypotheses based on your initial observations of the data

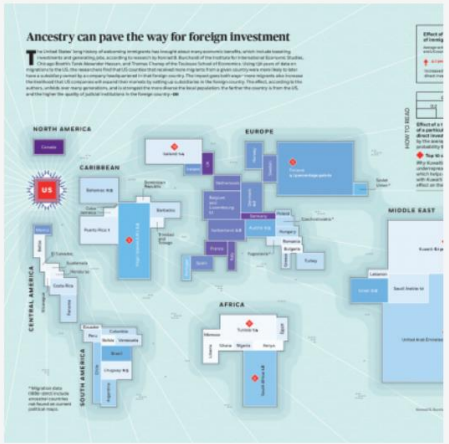
DATA VISUALIZATION

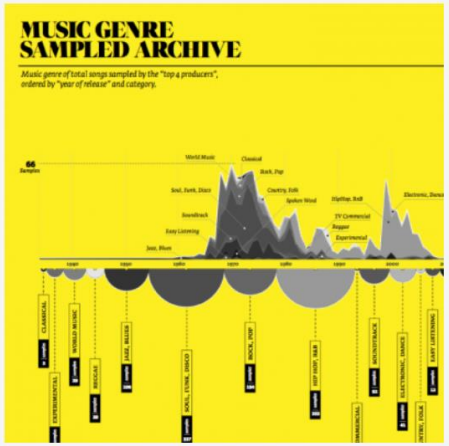


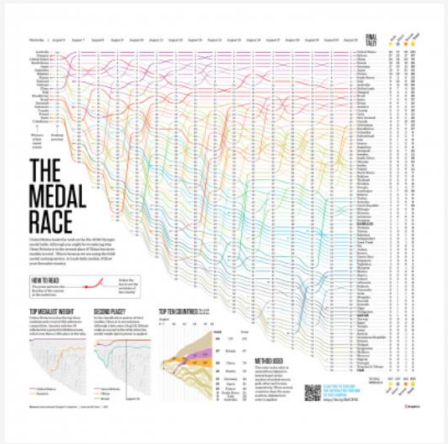












DATA VISUALIZATION

- **EXPLORE**: model your data in different ways
- **INTERPRET**: provide graphical evidence for trends
- **REPORT**: support arguments and convey information

PHASE 3: INTERPRETATION

Testing and Explanation

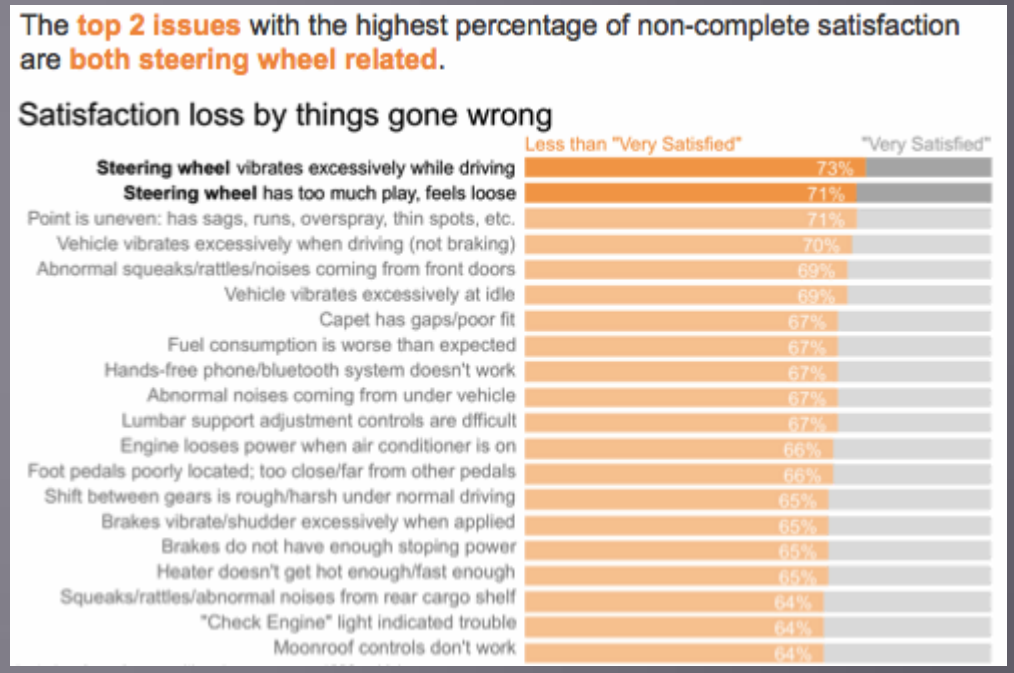
- Discipline-specific methodologies
- Matching expectations to the results of exploratory data analysis

PHASE 4: REPORT RESULTS

Communicating Your Results

- Clean versions of exploratory visualizations
- Storytelling with data

STORYTELLING WITH DATA



NYT example

THE FOUR PHASES OF DATA ANALYSIS

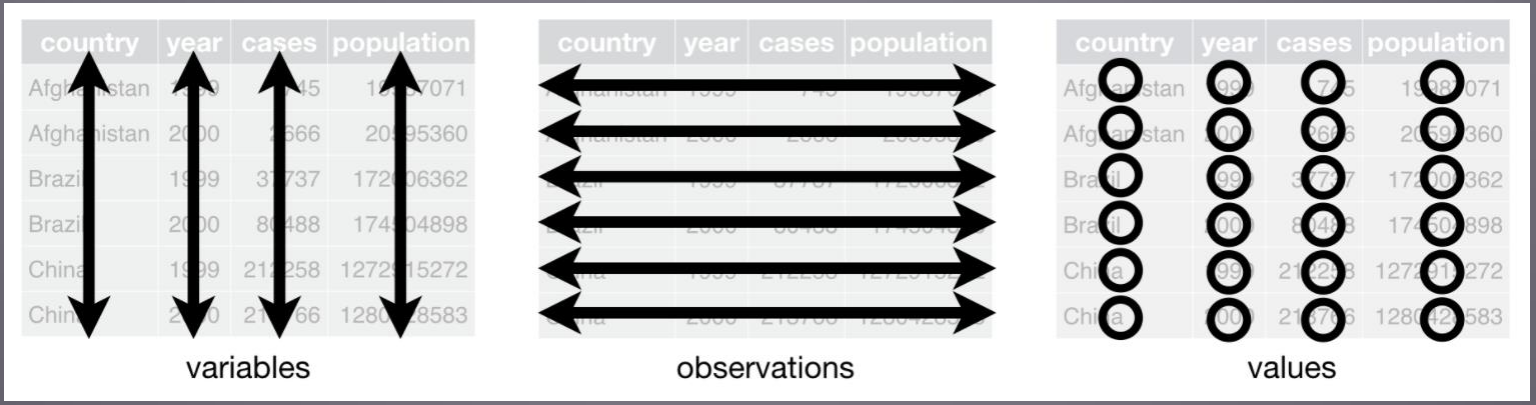
THE FOUR PHASES OF DATA ANALYSIS

DATA!

CLEAN DATA

Checklist:

1. Each variable in the dataset is placed in its own column
2. Each observation is placed in its own row
3. Each value is placed in its own cell
4. Missing values are all treated the same way
5. Duplicate observations are removed
6. Data is stored as an appropriate type



CLEAN DATA

Clean data is…

VALID	-	not empty; correct data type
ACCURATE	-	conforming to true values
COMPLETE	-	missing values treated properly
CONSISTENT	-	items across dataset match
UNIFORM	-	same units of measurement or text formats