Technology Fundamentals of Business Analytics

Visualization

Agenda

- Dave Geoghegan, Serial Entrepreneur
  - Chief Technology Officer of Channel Eyes

Exam Review

Visualization

- What do visualizations do?
- How do you do visualizations?
- Introduction to Tableau

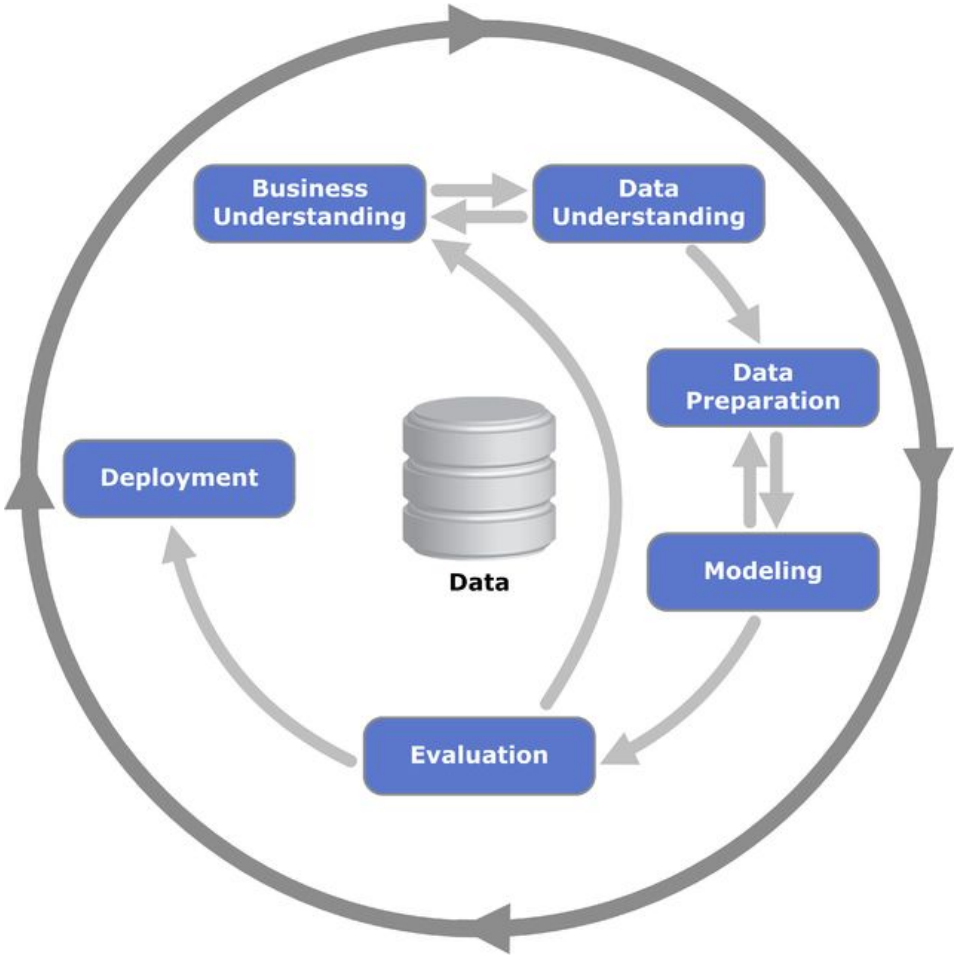
Exam Review

What do Visualizations do?

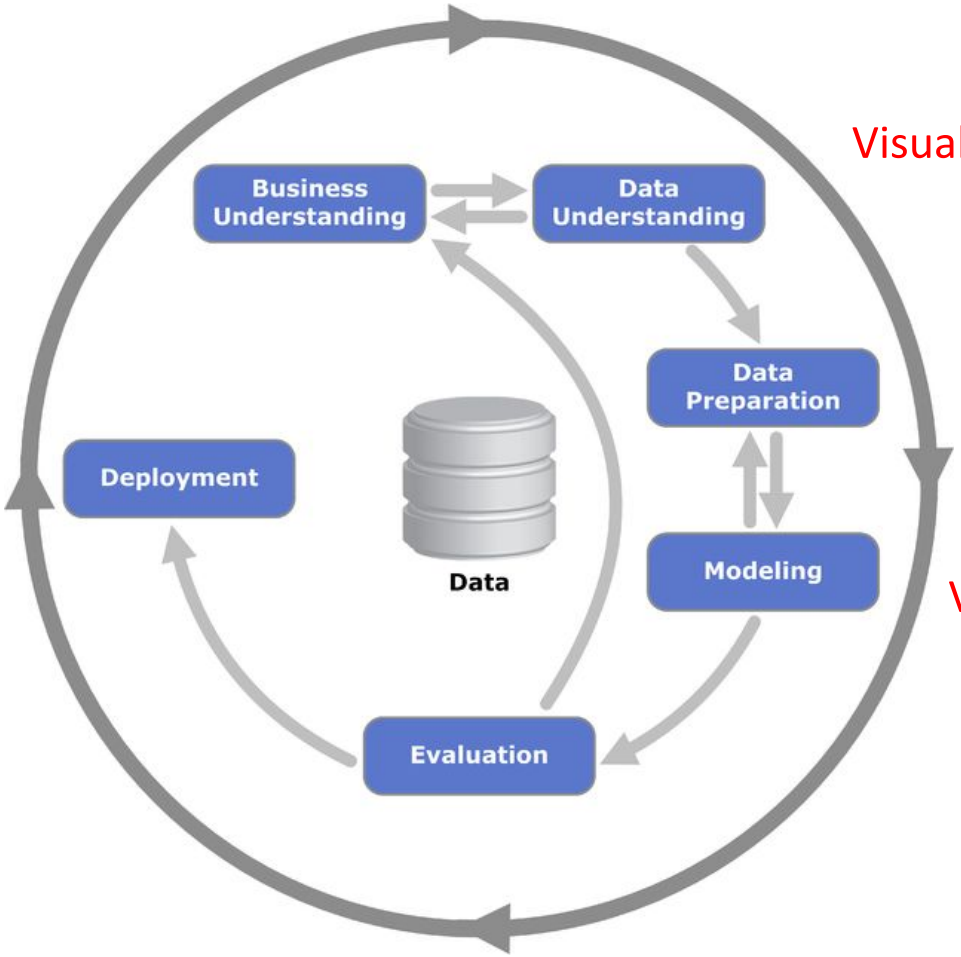
(1) Visualizations help the data scientist to understand the data and the relationships within the data

(2) Visualizations help the data scientist to communicate the data and the relationships within the data

Where does visualization happen?



Where does visualization happen?



Visualizations

Visualizations

Visualizations

Visualizations

Visualizations

Data Preparation

Filtering

- What is the data on which you would model? What is the range of the dataset? What type of filters should be used?

Transformation

- Should the variable be continuous or discrete?
  - **Example: {33.4, 45.4, 32.2, 32}**
- Log scale?

Feature Creation

- Create new Variables?

Modeling

Understand the Data

- What is the distribution?
- Are there any outliers?
- Do different features (groups/) relate to other groups

Communicate the Data

- Provide a reader with a visual to **more systematically understand data**
- **Tell a Story**

Data Understanding **Deployment**

Section/Modeling/Evaluation

Modeling

- Which features are relevant to the story?
- How should the data be represented?
  - Line, scatter, bar, colors shapes
- Can the data be easily interpreted as shown?

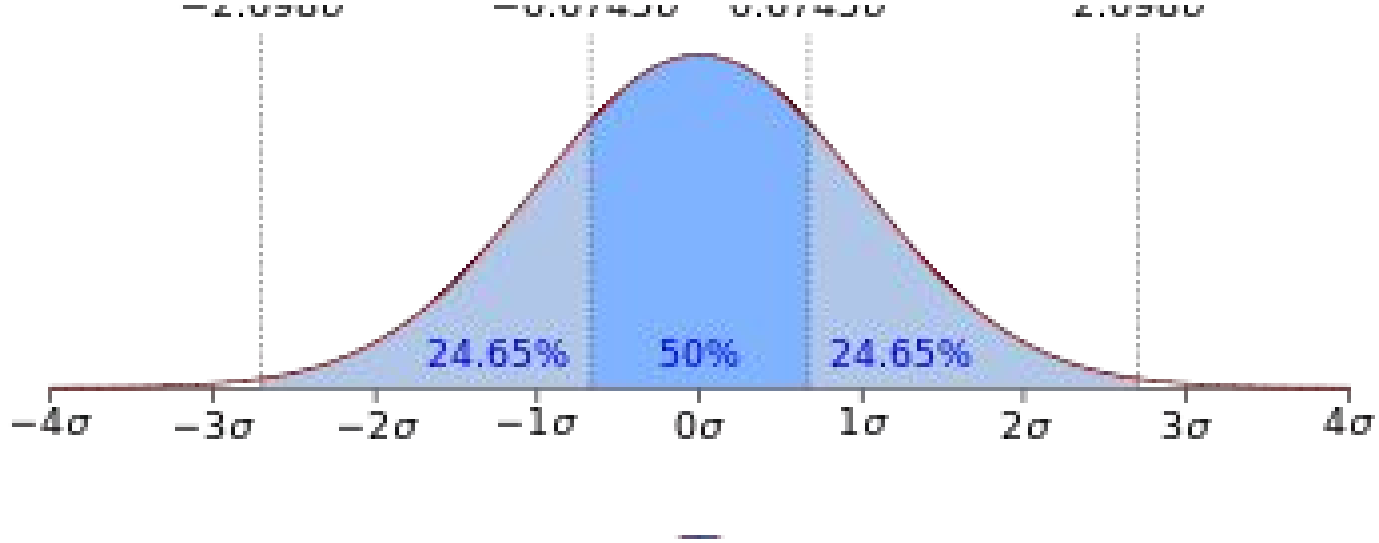
Understanding

- Does the visualization lead to new understand regarding data
- Does visualization facilitate understanding

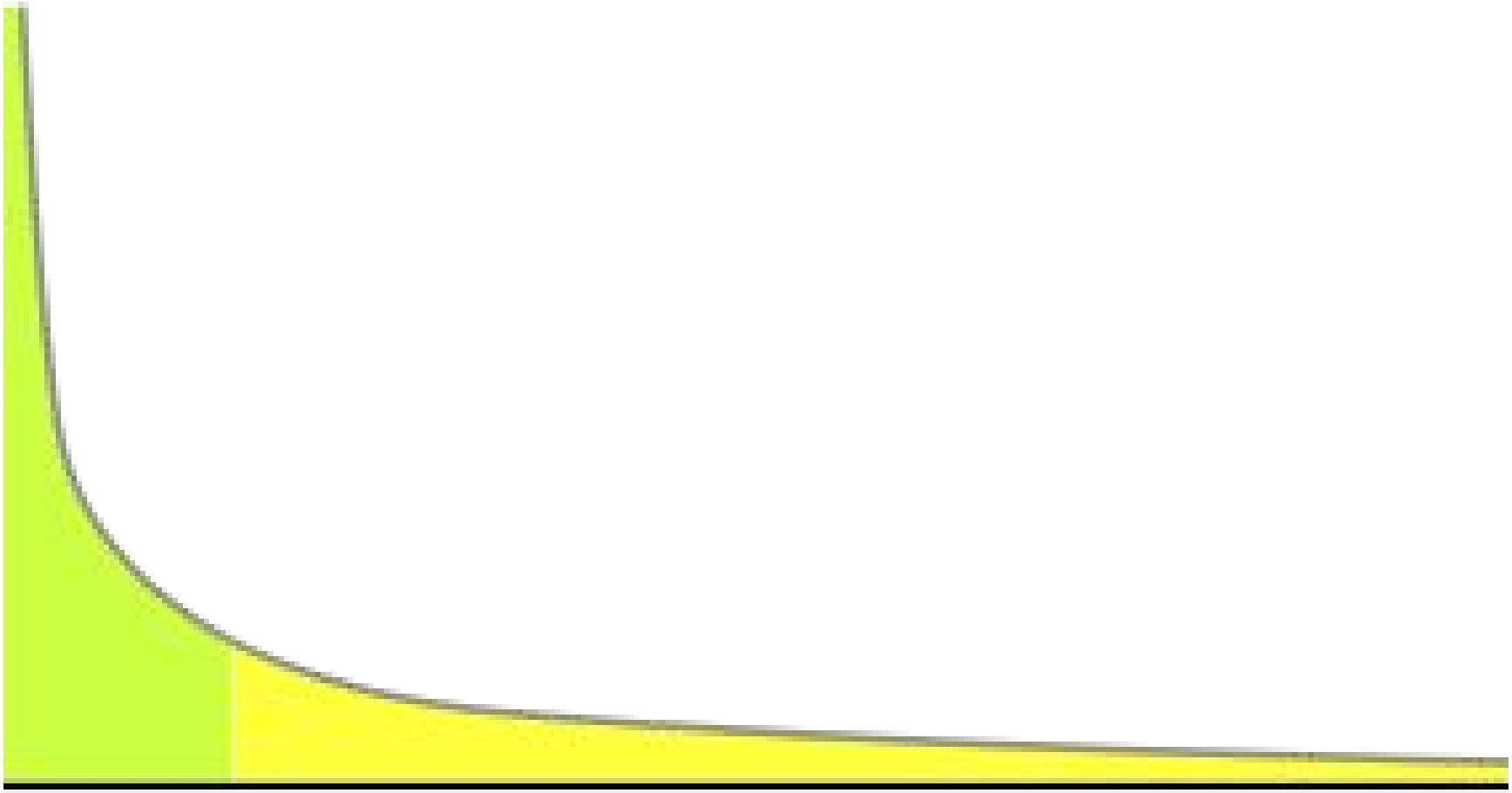
Distributions

- In probability and statistics, a probability distribution assigns a probability to each measurable subset of the possible outcomes of a random experiment, survey, or procedure of statistical inference.

Normal Distribution



Power Law Distribution



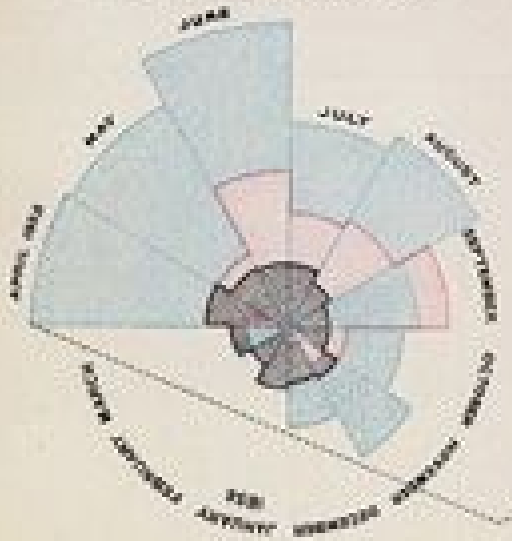
Distribution

- X-Axis (Columns): The value of the variable
- Y-Variable (Rows): Count of the number of rec at the value

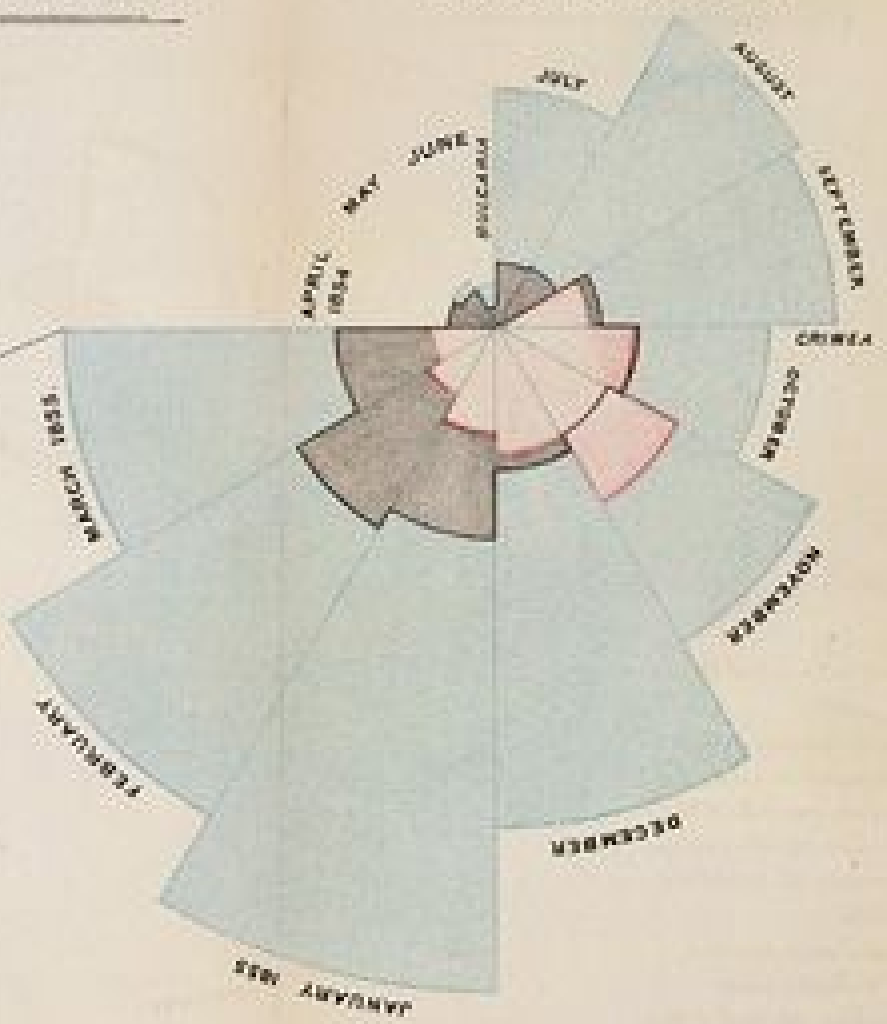
Examples

DIAGRAM OF THE CAUSES OF MORTALITY IN THE ARMY IN THE EAST.

2.
APRIL 1855 TO MARCH 1856



1.
APRIL 1854 TO MARCH 1855.



The Areas of the Blue red & black wedges are each measured from the centre as the common vertex.
The blue wedges measured from the centre of the circle represent area for area the deaths from Preventable or Mitigable Zymotic diseases, the red wedges measured from the centre the deaths from wounds, & the black wedges measured from the centre the deaths from all other causes.
The black line across the red triangle in Nov.r 1854 marks the boundary of the deaths from all other causes during the month.
In October 1854, & April 1855 the black area coincides with the red, in January & February 1855 the blue coincides with the black.
The entire areas may be compared by following the blue, the red & the black lines enclosing them.

Alternate Representations

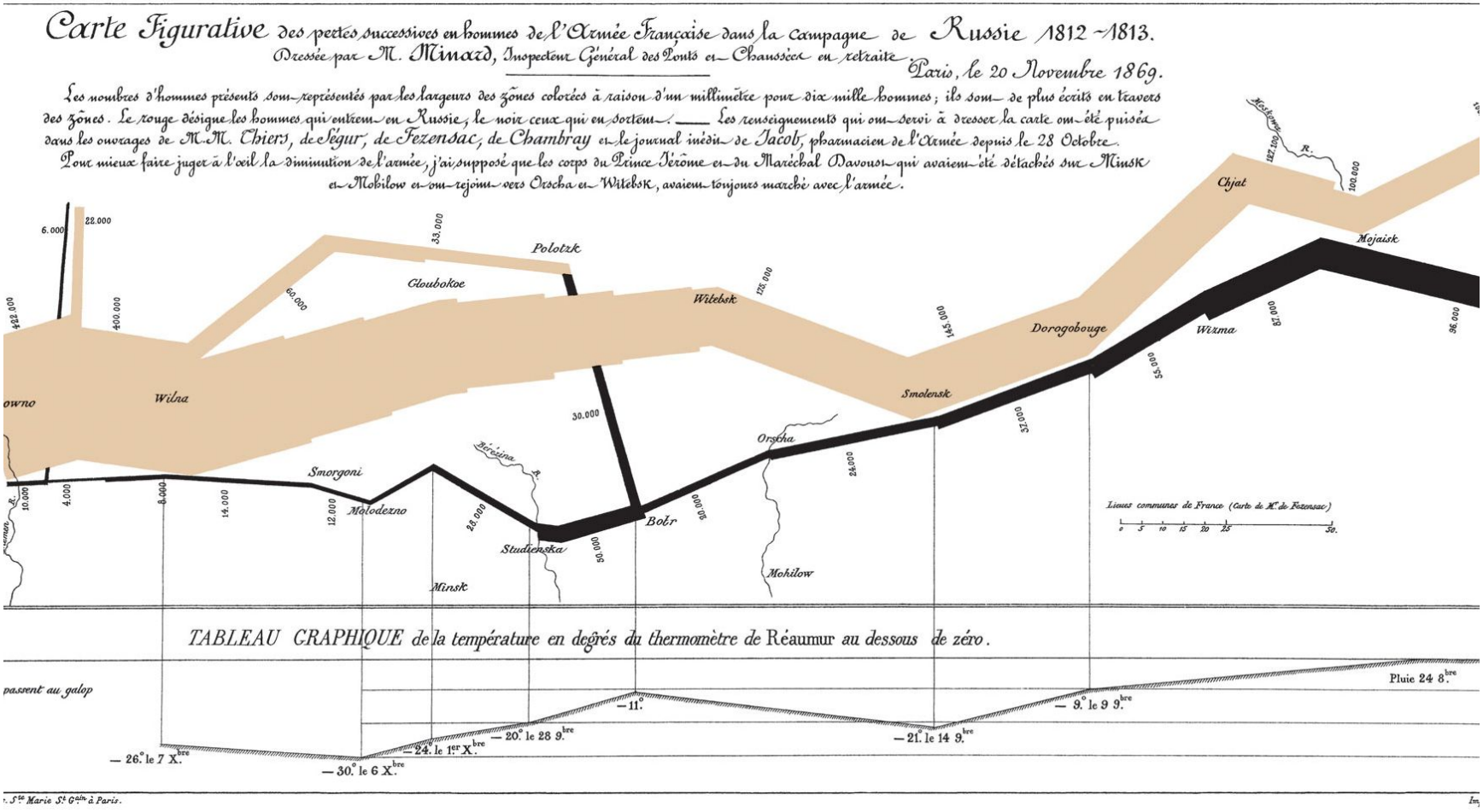
- http://understandinguncertainty.org/coxcombs

What Dimensions are shown?

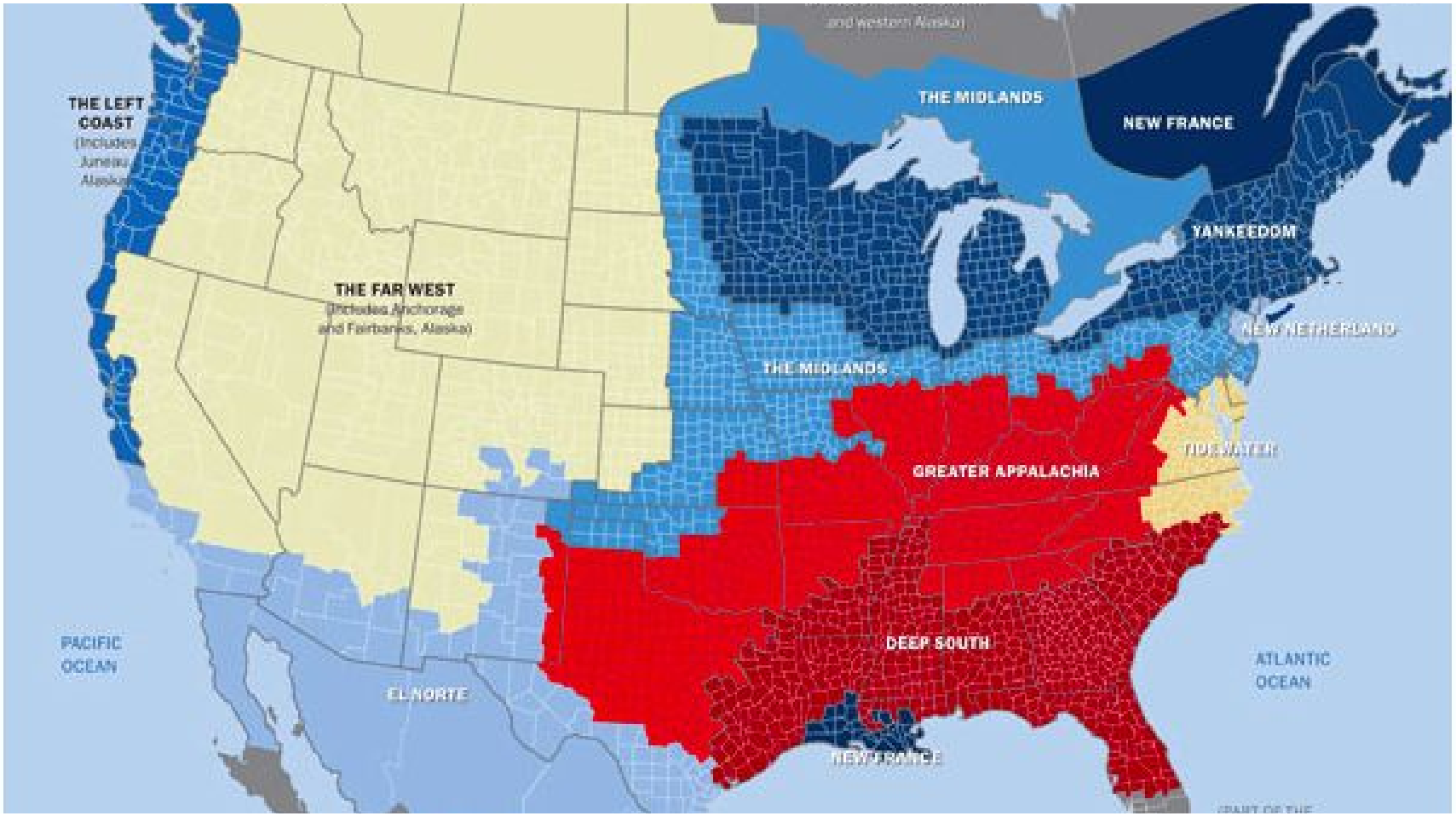
What Dimensions are shown

- Number – Number of casualties
- Cause of Death – What is the cause of death
- Month – When did the issue occur?

Napoleon's March to Moscow



Differences in Country



LIVE WIND MAP

http://hint.fm/wind/

Food Consumption

http://www.foodservicewarehouse.com/calorie-viz/

GAPMINDER for a fact-based world view

HOME | GAPMINDER WORLD | DATA | VIDEOS | DOWNLOADS | FOR TEACHERS | LABS | IGNORANCE

Explore the world

Gapminder World shows the World's most important trends

- Wealth & Health of Nations
- CO_2 emissions since 1820
- Africa is not a country!
- Is child mortality falling?
- Where is HIV decreasing?

Load Gapminder World

http://www.gapminder.org/

http://www.gapminder.org/videos/the-joy-of-stats/

How Might We Assign City labels based on Longitude/Latitude?

David McCandless:

The beauty of data visualization

TEDGlobal 2010 · 17:56 · **Filmed** Jul 2010

29 subtitle languages

View interactive transcript

Billion Dollar O'Gram

Presentation

http://www.informationisbeautiful.net/visualizations/billion-dollar-o-gram-2013/

Data

https://docs.google.com/spreadsheets/d/13qx5T10HFi0aT412MzoXZLjEljO78JX11AgljapzUug/edit#gid=13

Billion Dollar O'Gram

What is powerful in the visualization?

How do they accomplish this?

Billion Dollar O'Gram

- Label = Text Description
- Motivation = Colors = Category
- Amount = Size
  - Enables comparison of unrelated items
- Mouse Over
  - Long Description
- Click
  - Link to original source

Mountains Out of Molehills

Visualization

http://www.informationisbeautiful.net/visualizations/mountains-out-of-molehills/

Breakups on Facebook

Visualization

- http://www.informationisbeautiful.net/2010/peak-break-up-times-on-facebook/

Visually Normalizing Data

Visualization

http://www.theguardian.com/news/datablog/2010/apr/01/information-is-beautiful-military-spending

What is the importance of data understanding here?

Knowledge Compression

Visualization

- http://www.informationisbeautiful.net/visualizations/snake-oil-supplements/

What is knowledge compression?

Tableau

Example Data
(Describe – What do managers do)!

Tableau Overview

- Dimensions
  - Categorical Variables/Discrete
- Measures
  - Numeric Variables (Metrics)
  - Often Continuous

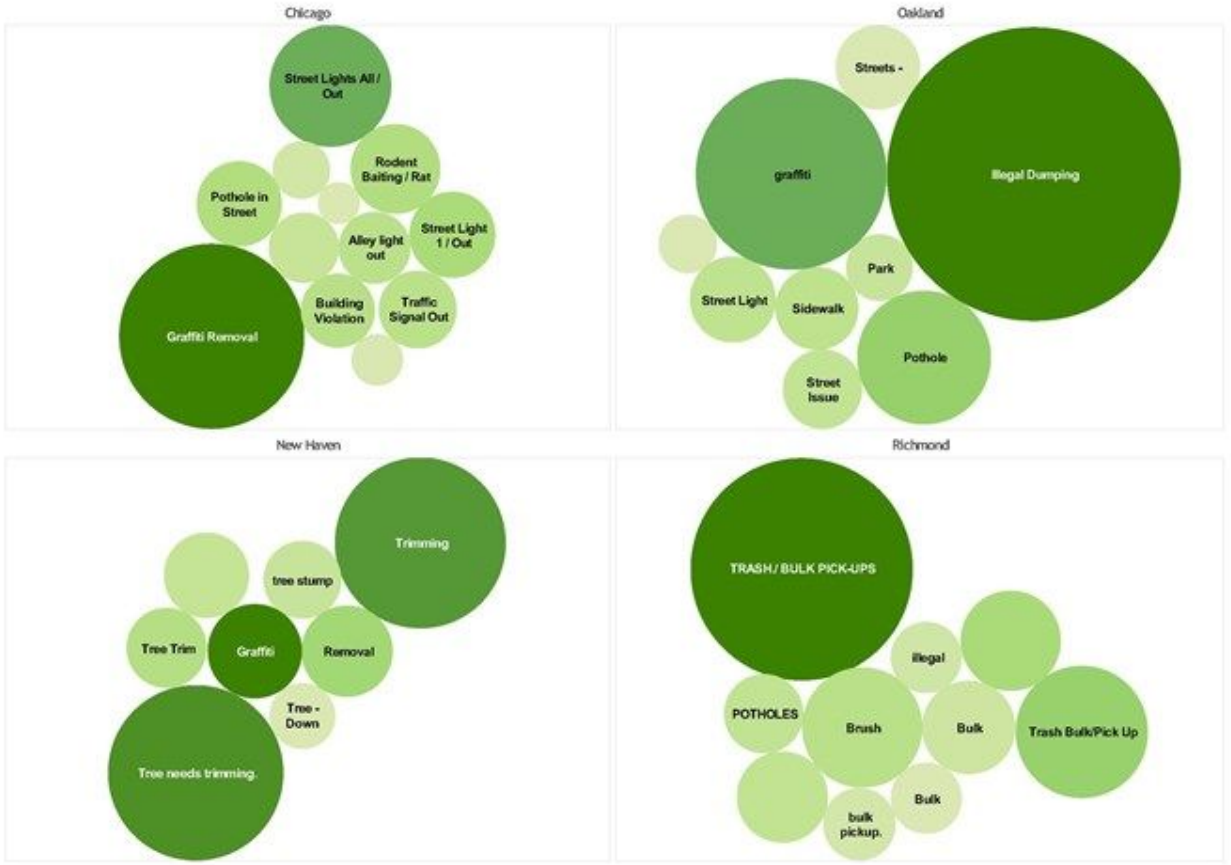
Various Aspects

- Show Me -> Hold Down Control and pick different chart types
- Filter (show quick filter)
- Plusses (minuses on views)
- Sort by axis
- Trend Line

See Predict Fix

https://www.kaggle.com/c/see-click-predict-fix

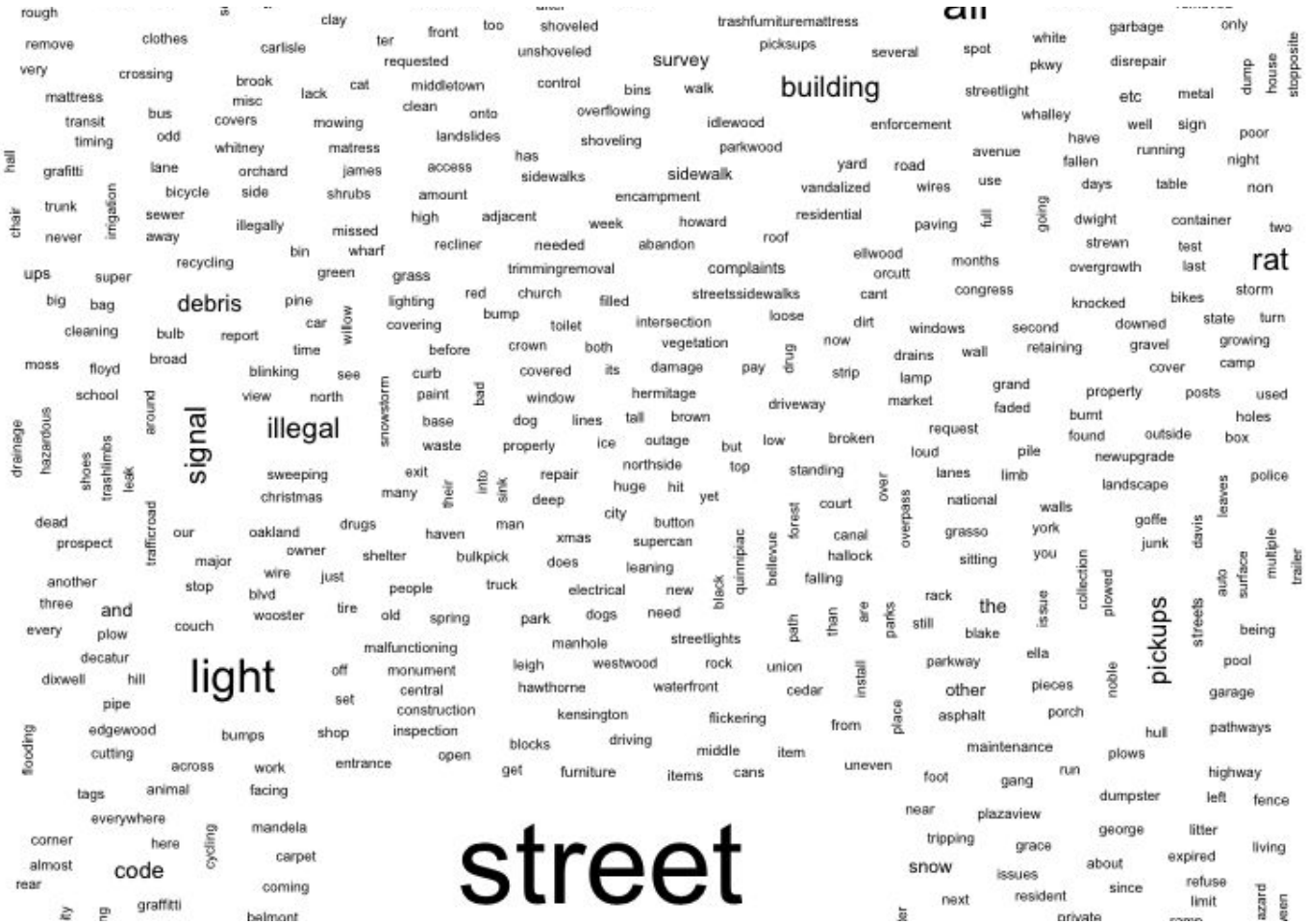
Visualization – See Predict Fix



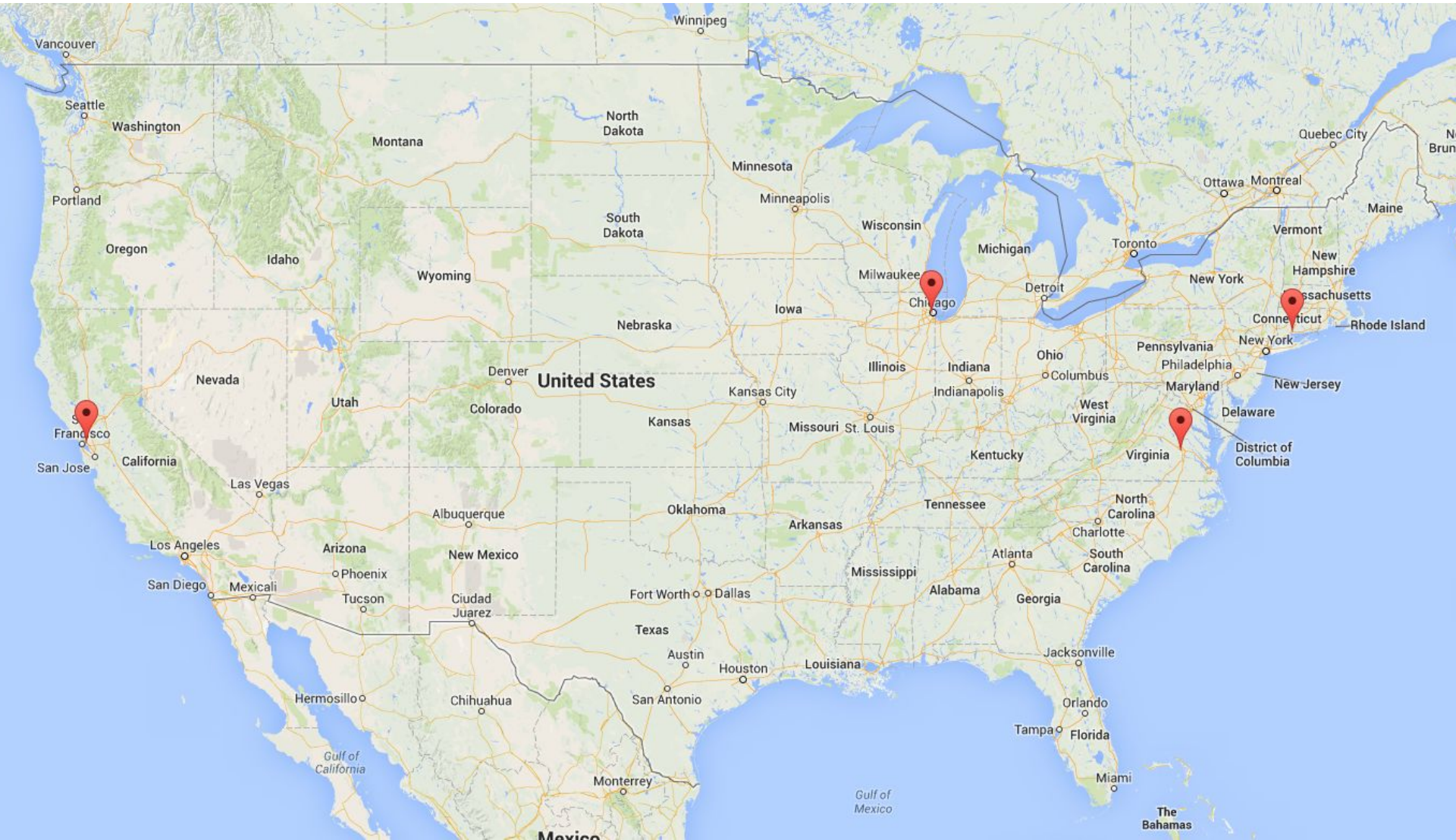
Visualization – See Predict Fix

SeeClickFix predict	Crime	Cars	Humans	Animals	Plants	Nuisance	Water
Oakland, CA							
Chicago, IL							
Richmond, VA							
New Haven, CT							
Tag Classes	Crime: graffiti, abandoned_vehicle(s), drug_dealing, robbery, illegal_idling, prostitution	Cars: pothole, street_light, signs,traffic, road_safety, bridge, parking_meter, bad_driving, street_signal	Humans: sidewalk, bike_concern, crosswalk,bench, pedestrian_light, homeless, lost_and_found, public_art	Animals: rodents, animal_problem, roadkill	Plants: tree, overgrowth	Nuisance: trash, blighted_property, noise_complaint, odor, heat', zoning, public_concern, test	Water: drain_problem, flood, hydrant, snow

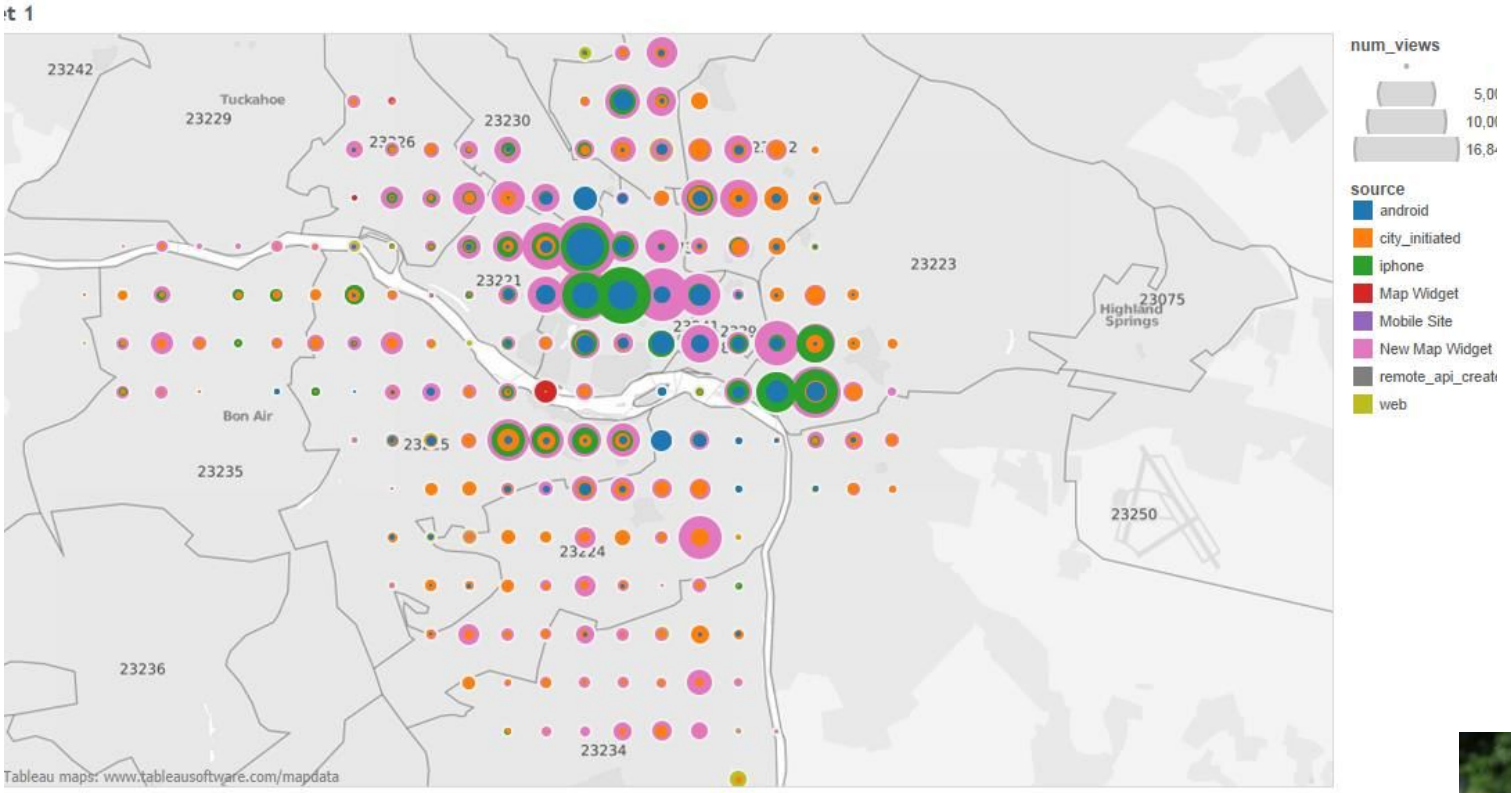
See Click Fix



Graph – D3 and Map



Kaggle Winner – See Predict Fix



ased on longitude and latitude. Color shows details about source. Size shows sum of num_views. The view is filtered on source, which excludes NA.

Anqi Wu, Business Analytics 2014