Introduction to Data Visualization Part 1: Developing Skills

Reed Jones

University at Buffalo

Hosted by ALCTS, Association for Library Collections and Technical Services

Overview

- Design: Things I wanted to know when I started working with data
- Chronological
- No SQL needed (but it is a good thing to learn)

What is Data Visualization?

- A way of arranging data in a way that makes it easy to understand its significance.
- Effective and Efficient
- Typically graphical

Get to know your data

- Plan ahead
  - Unique identifiers
  - Necessary fields
  - Goals
- Simplify data structures
- Check for validation issues
- Test, test, and re-test
- Get another set of eyes

Data Cleanup

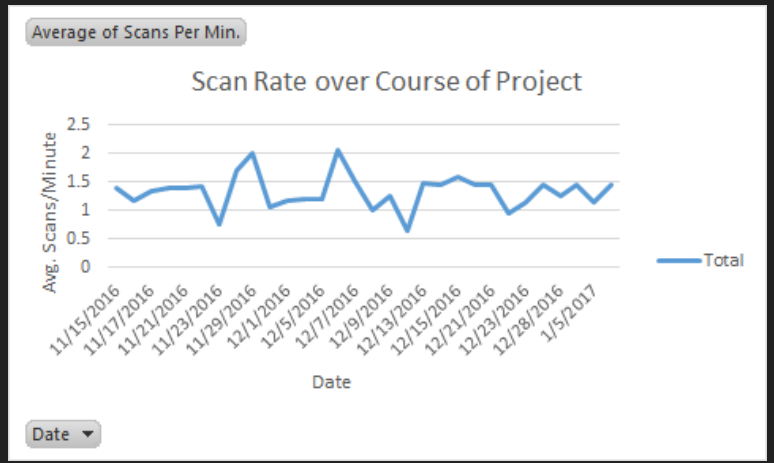
- Constant process
  - You can always go back to the data
- Anticipate where errors occur
  - Get in the mindset
- Create validation rules
  - Controlled vocabularies
- Develop a data entry workflow
- Communicate

Types of Visualizations

- Bar Graph
  - Comparing groups

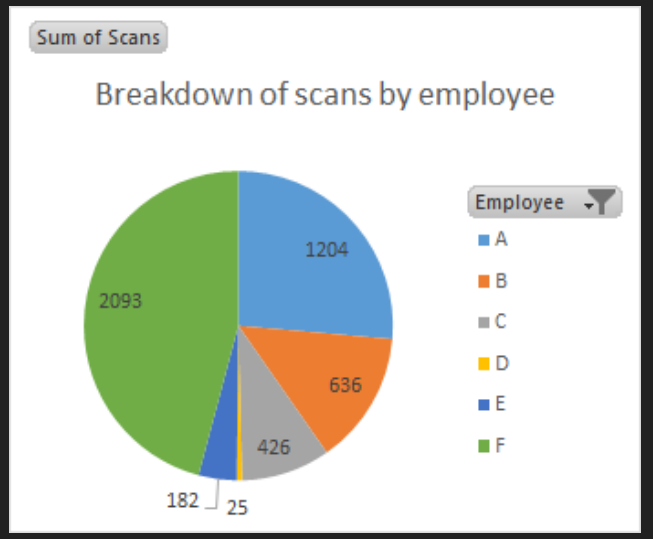
Types of Visualizations

- Line Graph
  - Change over time



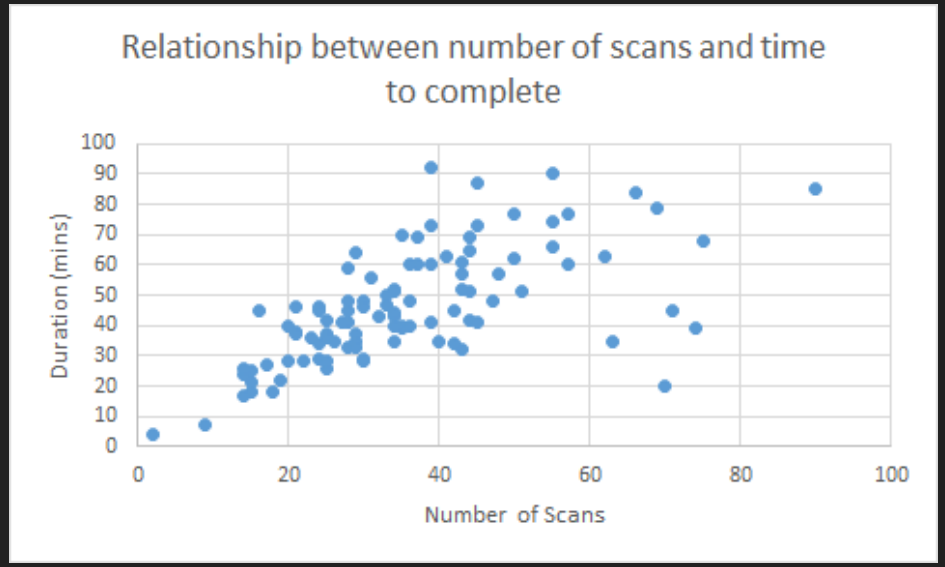
Types of Visualizations

- Pie Chart
  - Display parts of a whole



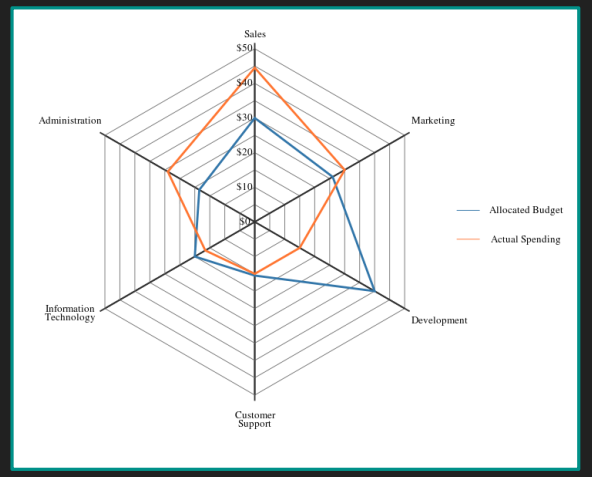
Types of Visualizations

- Scatter Plot
  - Relationships between two fields



Types of Visualizations

- Spider Graph/Radar Chart
  - Compare three or more fields
  - Great for relaying survey data



File:Spider Chart.svg, 2008, photo courtesy of Ordoon

Axis X/Y (Columns/Rows)

- Most visualizations only need a 2 dimensional representation
- General rule:
  - To get a visualization you need at least one dimension (element) to compare and one thing to measure
- The most frequent use of data visualizations have a dimension (element) on the x axis and a measure on the y axis

Measures vs. Dimensions

- Dimensions are useful in categorizing your data
  - Usually qualitative
  - Can sometimes be numerical
- Measures are calculations based on you data
- Dimensions can be turned into measures
  - Sum
  - Avg
  - Count
  - Count (Distinct)

Calculated Fields

- Sometimes we have to go back to the data
- Creating calculated fields sometimes makes life easier
  - Easy to identify the element and reuse
  - Visualizations process faster when using pre-made fields
- Common examples
  - Time difference
  - Fiscal Year

Filters

- Now it's time to dig in!
- Removes the noisy data
  - Contains
  - Equals
  - Is greater/less than
  - Begins with

Groups

- Overwhelmed by all the dimensions on your graph?
- Grouping fields together into different categories can make data easier to read.
- Ex: grouping academic departments into colleges

Sets

- Reusable filters that can be used on other visualizations
- Best if filtering one dimension at a time

Differentiation (Colors, Size, Shapes)

- Use these tools with purpose. (not simply an aesthetic choice)
- Complimentary to filtering
- Think in terms of user interaction

Labels and Keys

- Keep the title short and concise
- Each axis is labeled logically
- Use as little labeling as possible for your data
- Use unit labels (m, $, etc.)
- Utilize slicers/pages for the user to interact with

Let's Try an Example

- What is the scan rate of each employee?
  - Is it impacted by the type of scanner used?

Tips for When you get Stuck (Which will Definitely Happen)

- Ask, ask and ask some more
- Search for the answer online
  - Forums and video tutorials are great!
- Post questions on forums
- Take a break

Helpful Resources

- Mr. Excel
- Tableau Community
- For those adventurous enough to try SQL
- Datasets to play with

Questions