Doug Raffle

http://stat.wvu.edu/~draffle/
raffled@gmail.com | 724.518.7446

PORTFOLIO

This is a collection of projects I've worked on that I think are particularly cool. Whether created for class projects, formal research, or just for fun (and practice), I hope you find them as interesting as I did. Feel free to browse the projects on my Github page (http://www.github.com/raffled/), or take the guided tour of selected projects below. Direct links to the formatted documents (Shiny apps, knitted HTML, or PDF) are provided, as well as to the specific repositories containing the source code.

For a more interactive version that's being constantly updated, you can also check out the portfolio section of my website (http://stat.wvu.edu/~draffle/portfolio.html)

SENTIMENT ANALYSIS

Mylan Pharmaceuticals

While I was toying around with topics to analyze on Twitter, a friend of mine who works for Mylan pointed me towards the stories that were flying around about bidding wars between three generic pharmaceutical companies: Mylan, Perrigo, and Teva. Mylan is a pillar of the Morgantown community, so their being bought out would be huge news in Morgantown.

We pulled the most recent tweets with the hashtag #Mylan on April 24, 2015 and performed sentiment analysis to see what the Twittersphere had to say about the generics wars.

Analysis:	http://stat.wvu.edu/~draffle/portfolio/mylanSentiment.html
Source Code:	http://github.com/raffled/Sentiment/

USGS SEISMIC DATA

The USGS data set includes information about 20,000 global seismic events that occured from January 1, 2015 to April 15, 2015.

Interactive Map

First, I created an interactive map showing the location of the seismic events using RStudio's Shinypackage. Each point represents one event. Events are colored according to event type (earthquake, exposion, etc.) and scaled by magnitude. Because plotting all the events takes a while, the map is restricted to the 8,480 events which occured in the six weeks of March 1, 2015 - April 15, 2015.

Exploratory Data Analysis

This EDA is a basic exploration of the USGS data set, written in R Markdown, including numerical and graphical summaries of the data set.

Classification

First, I wanted to play around with some classification models, so I wanted to see if the three variable describing the seismic event could be used to predict the type of event. While not a particularly practical problem (if there's a large seismic even near me, I care more about the magnitude than what caused it), it is particularly challenging because of the nature of the data: 96.4% of the observations are earthquakes. I tuned several different models and compared their results.

Map:	http://raffle.shinyapps.io/seismic
EDA:	http://stat.wvu.edu/~draffle/portfolio/seismicEDA.html
Classification:	http://stat.wvu.edu/~draffle/portfolio/seismicModelling.html
Source Code:	http://github.com/raffled/seismic/