## *Ruchita Trivedi Project-1 STAT-6610*

## 1. Reproduce the **More Students Are Earning** visualization that was posted to the White House twitter account  Search twitter for "WhiteHouse More Students Are Earning" and click on View Conversation.  Here is the link to the result [https://twitter.com/search](https://twitter.com/WhiteHouse/status/558318033639198720).

## Start by creating an Excel spreadsheet containing the data.

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |
|  | Years | graduating percent | |
|  | 2007-08 | 75 |  |
|  | 2008-09 | 75 |  |
|  | 2009-10 | 78 |  |
|  | 2010-11 | 80 |  |
|  | 2011-12 | 81 |  |
|  |  |  |  |

## What are the variables being presented?

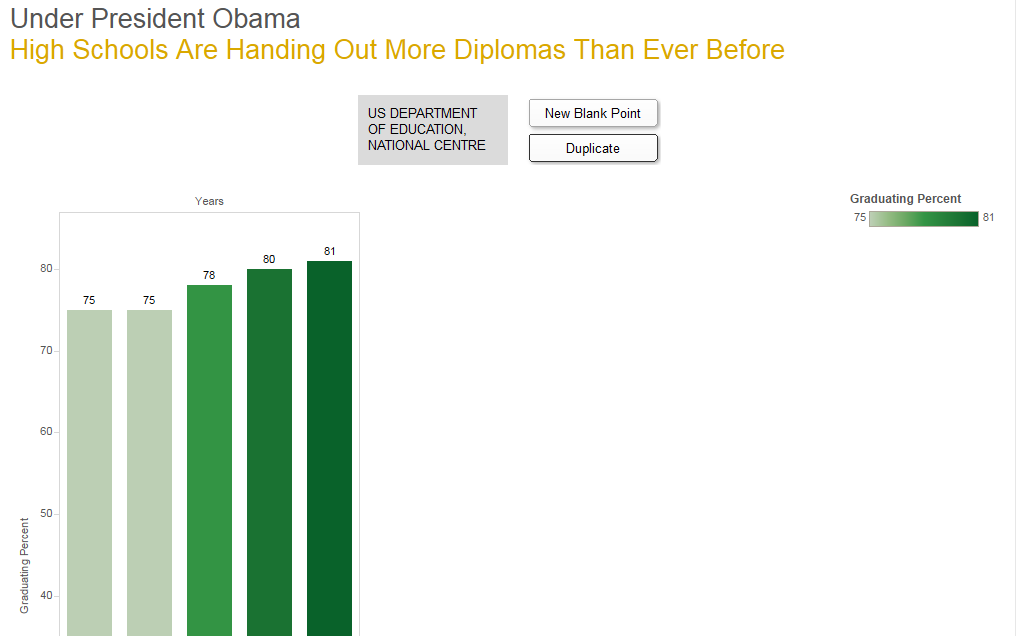
Years , Graduating Percent

## Name the Title and Source.

Title: High School Graduation Rate

Source: US department of Education, National Centre of Education Statistics

## Use Tableau to make the visualization.  It may be helpful to use a Story in Tableau.



## Explain what is wrong with the plot and how you have corrected the problem.

The plot showed exaggerated y-axis ie the graduation percentage, to give an effect that more and more students were earning diplomas with each passing year

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

2. Make the following plots:

* Make a bubble plot for a dataset that included the letters A and B, where the letter A occurs 100 times and the B occurs 50 times.
* Start by creating an Excel spreadsheet containing the data.

|  |  |  |
| --- | --- | --- |
|  |  |  |
|  | Words | Frequency |
|  | A | 100 |
|  | B | 50 |

* Use Tableau or R to make the visualization.



* Verify that the size of the circles are correct.

Since we labeled the frequency in the bubbles, we see that sizes are corresponding to frequency of words. The size of bubble A is twice that of B.

* Make a Word Cloud for a dataset that included the letters C and D, where the letter C occurs 300 times and the D occurs 400 times.

Start by creating an Excel spreadsheet containing the data.

|  |  |
| --- | --- |
| Word | Frequency |
| C | 300 |
| D | 400 |

* Use Tableau or R to make the visualization.



* Verify that the size of the circles are correct.

When we hover the mouse over image we can see the Word and Frequency , which verifies that the size s are correct. Also visually , C is three fourth the size of D.

3. The following [article and plot](http://www.sci.csueastbay.edu/~esuess/classes/Statistics_6610/Homework/Project1/labor.jpg) appeared in the Wall Street Journael on Tuesday, Feburary 10, 2015.

* Use the [China Labour Bulletin](http://www.clb.org.hk/en/) [Strike Map](http://maps.clb.org.hk/strikes/en) website.
* Start by creating an Excel spreadsheet containing the data for each year there is data available for Beijing, Guangdong, Shanghai, and Chonging.

[Attached the Spreadsheet]

* Make a time plot for each location using the same scales for the axes.  (Use Tableau or R)



* Make one plot with all four locations on the same time plot in different colors.  Use colorbrewer.  (Use Tableau or R)



* Make a dynamic bubble graph using Tableau.







* What is a possible problem with the data selected for comparison.  (Hint: Think about city populations size.  See the [City Population](http://www.citypopulation.de/) website.)

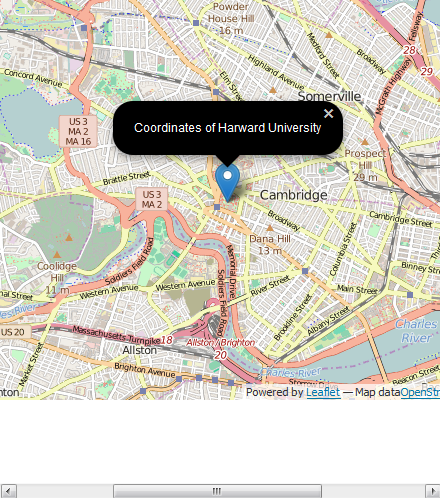
|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | 2015-08 | 2015-09 | 2015-10 | 2015-11 | 2015-12 | 2016-01 | 2016-02 |
| Beijing | 3 | 9 | 4 | 7 | 15 | 6 | 1 |
| Guangdong | 32 | 29 | 52 | 56 | 56 | 53 | 3 |
| Shanghai | 5 | 6 | 3 | 7 | 7 | 11 | 0 |
| Chonging | 4 | 4 | 2 | 7 | 9 | 8 | 0 |

We see that Guangdong has largest radius than other bubbles, so we look into the labor strikes

By states and by year and see that Guangdong had the highest numbers for all years.

The above table shows that numbers for Guangdong are high, ie the population density may be more or the state may be bigger in area.

4. Try the code from [HelloFromCSUEastBay](http://rpubs.com/esuess/HelloFromCSUEastBay).  Change the location and the popup.  What is leaflets?  Show the plots you have created.



Leaflet is the leading open-source JavaScript library for mobile-friendly interactive maps. Weighing just about 33 KB of JS, it has all the mapping [features](http://leafletjs.com/#features) most developers ever need.

Leaflet is designed with simplicity, performance and usability in mind. It works efficiently across all major desktop and mobile platforms, can be extended with lots of [plugins](http://leafletjs.com/plugins.html), has a beautiful, easy to use and [well-documented API](http://leafletjs.com/reference.html) and a simple, readable [source code](https://github.com/Leaflet/Leaflet) that is a joy to [contribute](https://github.com/Leaflet/Leaflet/blob/master/CONTRIBUTING.md) to.

5. Try the code from [WeatherToday](http://rpubs.com/esuess/WeatherToday).  (Updated 2/23/2016 removed LAX, compare Seattle with SF.) The R package weathermetrics connects to Weather Underground data.  Change the aiports and rerun the code.  Show the plots you have created.

Note: Changed the Airports to Las Vegas and Denver for the comparison.

