

GEORGE MASON UNIVERSITY

STAT 515

REDESIGN PROJECT REPORT

GROUP 10

BY

ABHISHEK SHAMBHU

AAKANKSHA AUNDHKAR

JINYI WU

25-10-2018

Table of Contents

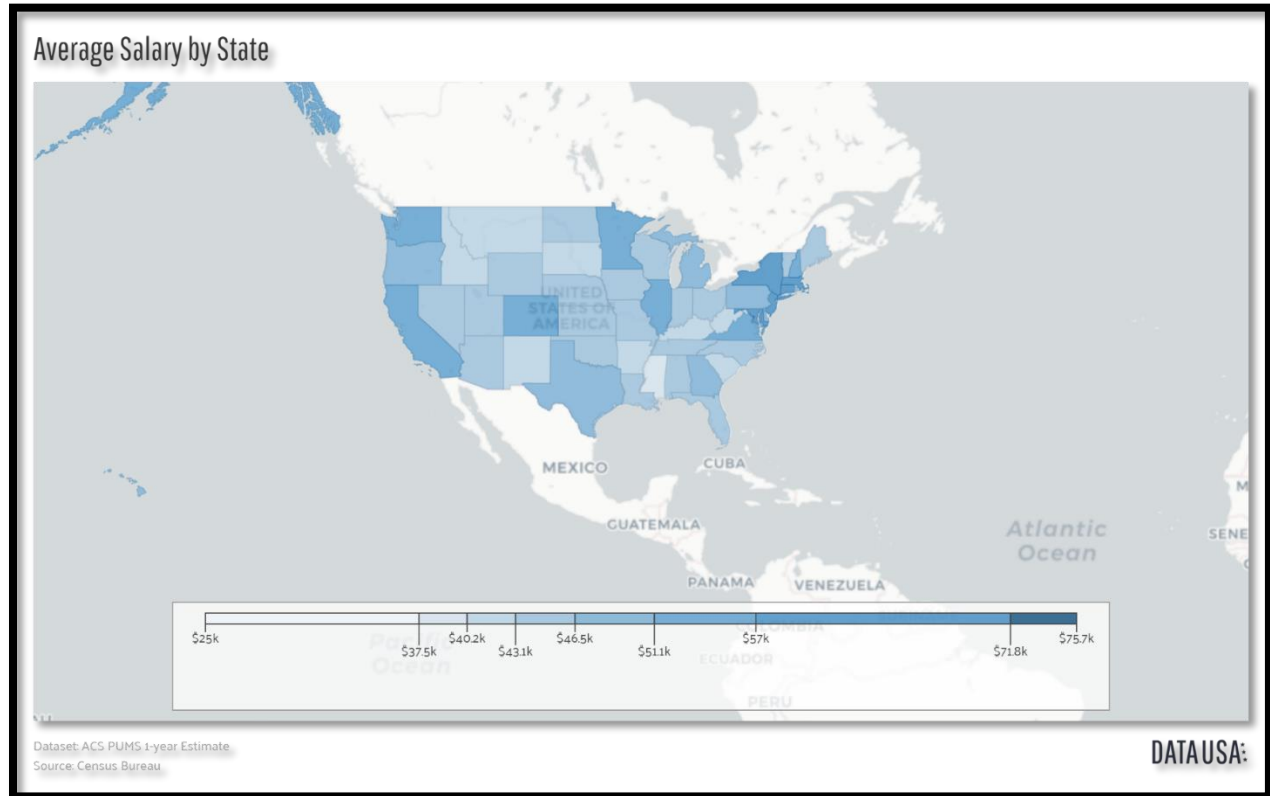
1. Introduction
2. Graph 1
 - a. Original Graph
 - i. Description
 - ii. Design Flaws
 - b. Redesigned Graph
 - i. Description
 - ii. Design Improvements
3. Graph 2
 - a. Original Graph
 - i. Description
 - ii. Design Flaws
 - b. Redesigned Graph
 - i. Description
 - ii. Design Improvements
4. Graph 3
 - a. Original Graph
 - i. Description
 - ii. Design Flaws
 - b. Redesigned Graph
 - i. Description
 - ii. Design Improvements
5. References

Introduction:

The main objective of doing a redesign project is to make charts and graphs easier to understand so that it can interpret information quickly and accurately. Graph such as pie charts, area charts are harder to understand and to say which area is bigger or small than the other. A 3D graph for different measures but no projections to the axis doesn't convey much and is harder to understand. A 2D graph in this case will give us a better idea about the value for the parameters and will tell us which is taller or shorter than the other. Scatterplots are also an effective visualization for showing relationships between the two variables plotted on the x and y axes. Use of different perceptual groupings such as color, size, orientation, etc. will help us to analyse and visualize the data in a better way.

Graph 1

Original Graph:



(Source: Data taken from <https://datausa.io/map/>)

Description:

The dataset for the above graph consists of 50 states of USA and the average salary with respect to each state. Moreover, this website has other parameters too for the 50 states such as Average Age per state, Population per state, Average Income per state, etc. So, to work on a dataset for redesigning we have merged the parameters for the 50 states into one single dataset. Also, I have added the GDP column into my dataset for the 50 states (taken from Wikipedia) which I have used further in my redesigning graph.

Design Flaws:

The main problem with the above graph is that it only describes the map based on a single parameter. In the graph above, it can be seen that it is just showing the Average Salary for each state. Also, the graph has a slider at the bottom which gives us the range in which a certain State average salary lies in, but it would have been better if it would have used a label to show value for each state or a better representation to more accurately identify the average salary. Moreover, it can be seen that the graph has a monochromatic color representation with different shades for representing different range for the average salary. In the above graph, it can be seen that the darkest blue is a state with the highest average salary whereas the light blue color at the left of the slider shows the lowest average salary for a state. It would have been better if the graph could have used different colors for showing states with different average salary ranges. In addition, the graph since it uses a single parameter to show patterns it is not possible to establish relationship without addition of other parameters in the data and so needed to merge with other data parameters so as to work on redesigning the graph. So, for simplicity and better representation of the above graph we would redesign the above with Linked Micromaps.

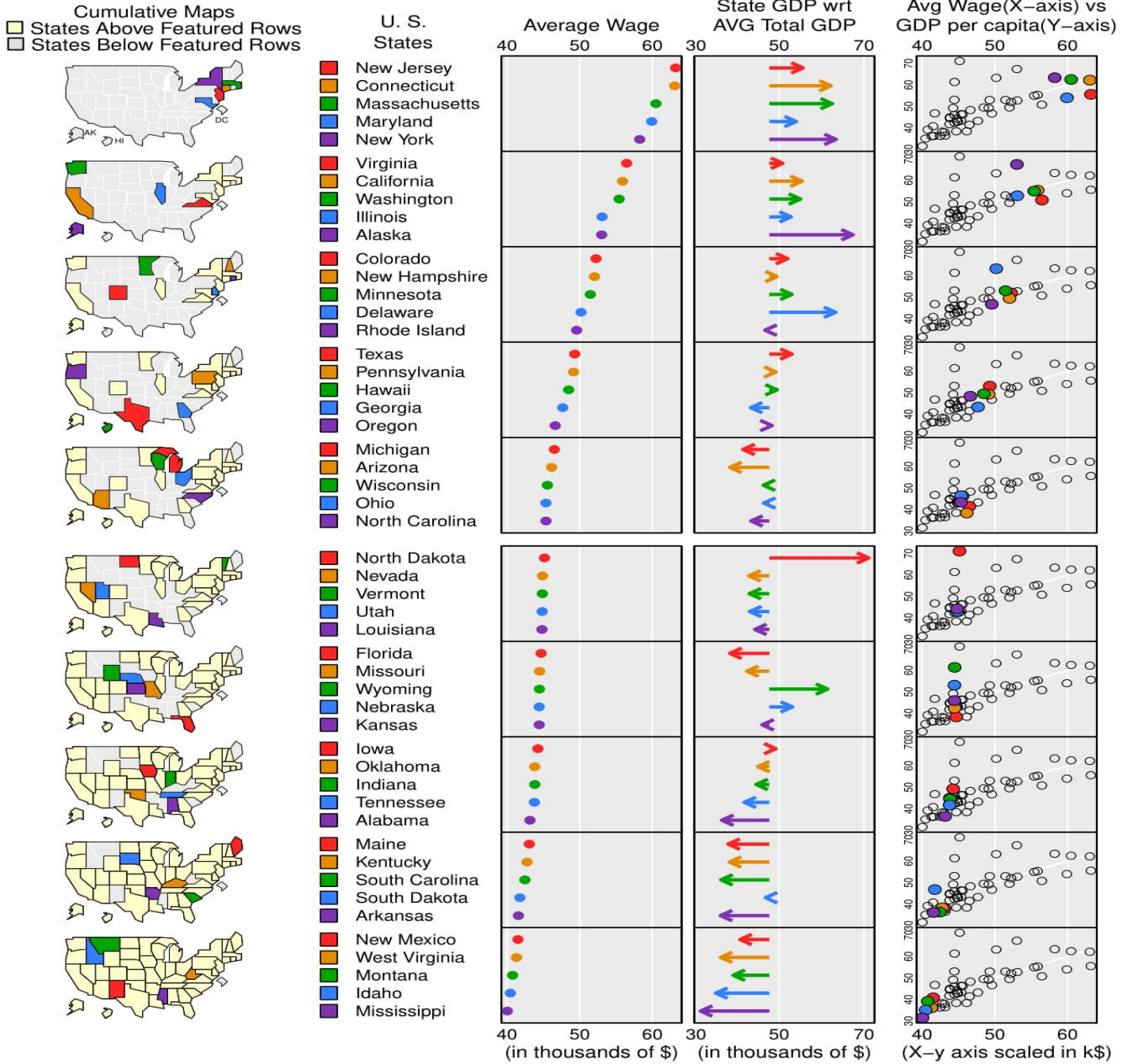
Table below shows the first few rows of the dataset built by merging multiple data set parameters. Here, we are only using the average wage and the GDP column which is scaled in our redesign graph.

(Note: The GDP in the last column is scaled and is the GDP per capita value.)

State	age	pop	Ori avg wage	avg wage	avg wage moe	Total wage	Ori GDP	GDP
Alabama	38.6	4841164	43170.8	43.17	658.283	2.09E+11	36538	36.5
Alaska	33.6	736855	53027.4	53.03	1886.44	39073504827	67411	67.4
Arizona	37.1	6728577	46143.6	46.14	532.131	3.10E+11	38427	38.4
Arkansas	37.7	2968472	41581.3	41.58	727.027	1.23E+11	36249	36.2
California	36	38654206	55901.3	55.90	318.719	2.16E+12	55374	55.4
Colorado	36.4	5359295	52249.3	52.25	626.704	2.80E+11	52019	52.0
Connecticut	40.6	3588570	63060.4	63.06	1018.79	2.26E+11	62236	62.2
Delaware	39.6	934695	50186.4	50.19	1601.53	46908977148	63271	63.3
Florida	41.6	19934451	44699.1	44.70	355.088	8.91E+11	38398	38.4
Georgia	36.2	10099320	47679.8	47.68	569.312	4.82E+11	43313	43.3
Hawaii	38.5	1413673	48476.1	48.48	1006.58	68529353715	49497	49.5
Idaho	35.7	1635483	40474.8	40.47	965.566	66195847328	35099	35.1

Redesigned Graph:

GDP PER CAPITA VS AVERAGE WAGE PER STATE IN USA IN 2016



Description:

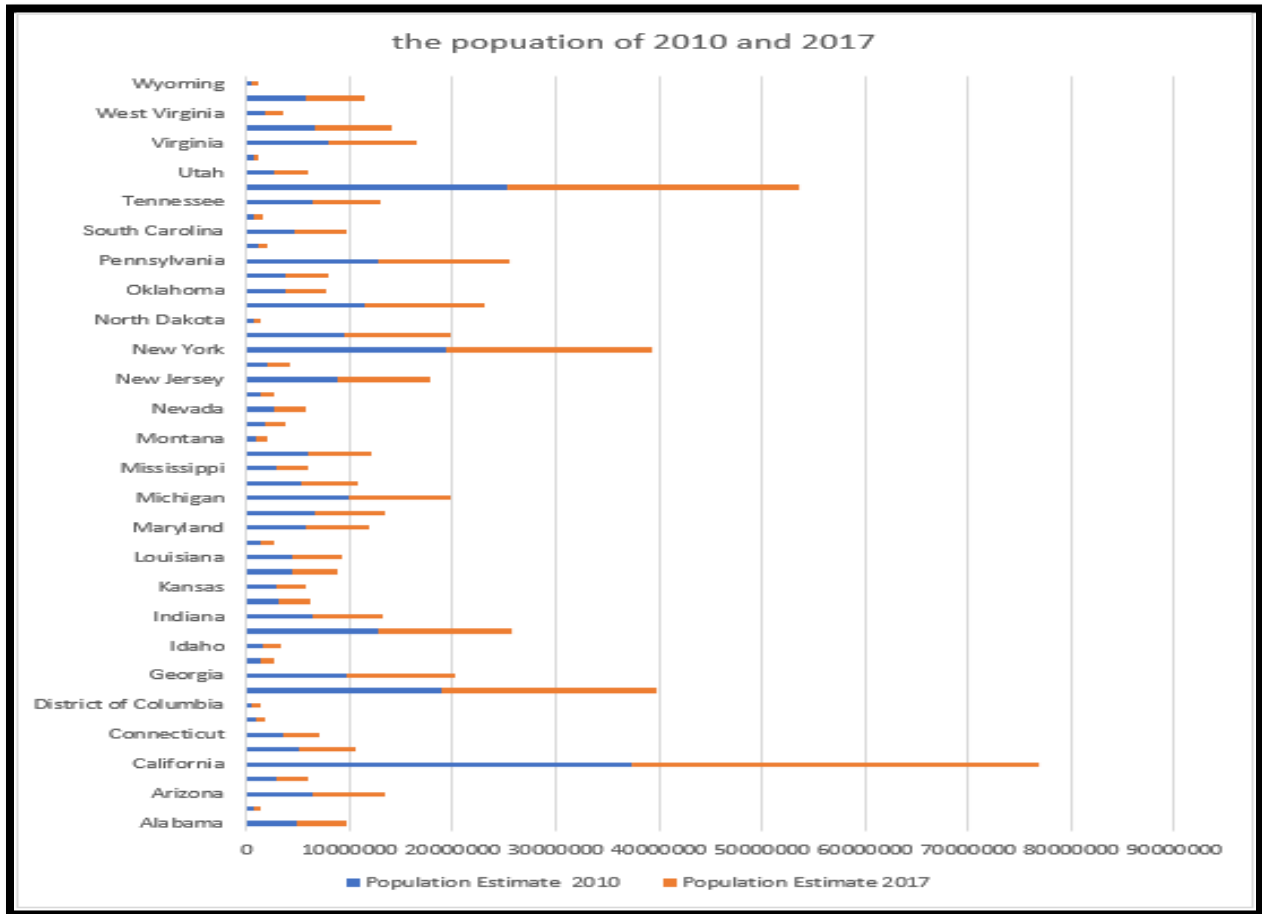
To redesign the Original Graph, I have used the 'micromapST' function to create Linked Micromaps which can be used to better visualize the data. Linked Micromaps are better since it not only adds small maps for each perceptual group of rows, but it also describes the graph in more detail. Since, I had the data for the 50 U.S. states in terms of Average Wage, GDP, Population, etc. I have used this column to better represent and show relationship in my data. Since, the data didn't have a GDP per capita column, I had used the GDP and the population data to get the GDP per capita for each state. Moreover, since the values were in thousands of dollars, I have rescaled the Average salary and GDP per capita so that it can be better in terms of visualizing the data.

Design Improvements:

In the above redesign graph, as it shows the first is the dot plot to plot the average salary per state and is sorted for states with the highest to the lowest average salary. The states in the map is highlighted based on the average salary. Moreover, I have used a cumulative map function to represent state groups with higher and lower average salary than the other perceptually grouped rows. Also, the U.S. state names has been mentioned in full for better understanding. The second is the state GDP per capita with respect to the total average GDP per capita where I had used arrows to represent states which have a higher and lower GDP per capita that the average. The third is a scatter plot of the Average Wage per state versus the GDP per capita to see if there are any relationships or correlation between the two parameters and it can be seen from the Line of Best fit that the two parameters shows a positive correlation. i.e. as the average salary goes from lower to higher, the GDP per capita also goes from the low to high. Also, in this graph they have mentioned labels and values for x and y axis to better identify value for each state. In addition, in this graph, we are using different colors to better visualize state data for different parameters instead of a monochromatic color with its different shades. It also gives us a clear understanding what each color is i.e. which color belongs to which state. Since, we have merged multiple datasets we are now able to find if there any relationships which was not possible in the original graph.

Graph 2

Original Graph



(Figure: Population estimate of each U.S. state in 2010 and 2017)

Description

In this project, the author hopes to compare the population changes of each state in the United States from 2010 to 2017 through a graphic as above. This graph depicts the population of each state in 2010 and 2017. We can distinguish between 2010 and 2017 populations in two different colors. From the ordinate is each state of the United States, the abscissa is the number of people. In this picture we can very intuitively see which state's population is ranked among the top states in the United States. For example, the population of California is significantly higher between states.

Design Flaws

The following points are some questions about the original graph as follow:

1. There are many states in the United States, so the visual effects don't look so clear.
2. Each bar shows the value for two years, which doesn't seem so easy.
3. This picture cannot compare the changes between the two years. For example, the change in growth rate or the amount of growth.
4. The vision is not strong and can't attract attention.

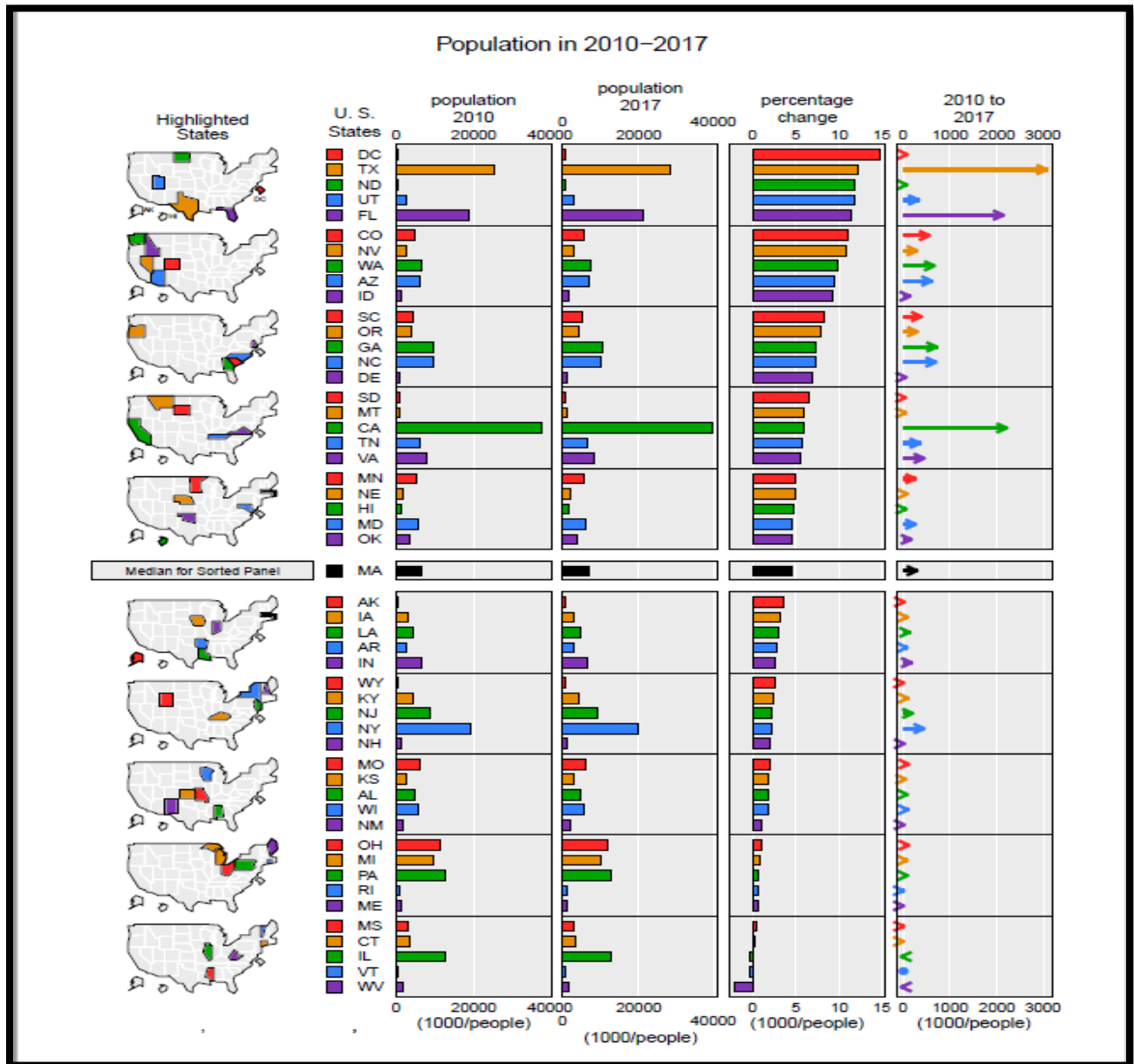
These are the questions I listed about this picture, I think I can redesign a picture using Micromap. In response to the problem I pointed out above, there are a few things that need to be done to improve the visualization. Firstly, because it is necessary to compare the changes in the US population from 2010 to 2017, it is necessary to increase the comparison of variables in the figure. Secondly, Because the demographic data is huge, we need to make the data smaller so that we can better compare and increase the visualization. Later I will explain how I made the data smaller.

Redesigned Graph

Description

About the redesigned map, I used Micromap on RStudio. Micromap has many advantages, such as the ability to display a variety of variables in a graph, and also display a mini map in the map, and the color is rich. This will help you analyse many, many variables in a single graph.

The below is my redesigned graph:



From this picture we can clearly obtain data on demographic changes. First of all, we can view the location of each different state through the mini map, and it is presented in color, which makes it easy to distinguish. However, we see that the bar data in the first two columns shows the population of each state in 2010 and the population of each state in 2017. Finally, we are able to achieve growth rates in the “percentage change”, positive growth and negative growth from 2010 to 2017 for the state population. Then we can see the arrows that can indicate the number of population growth changes from 2010 to 2017.

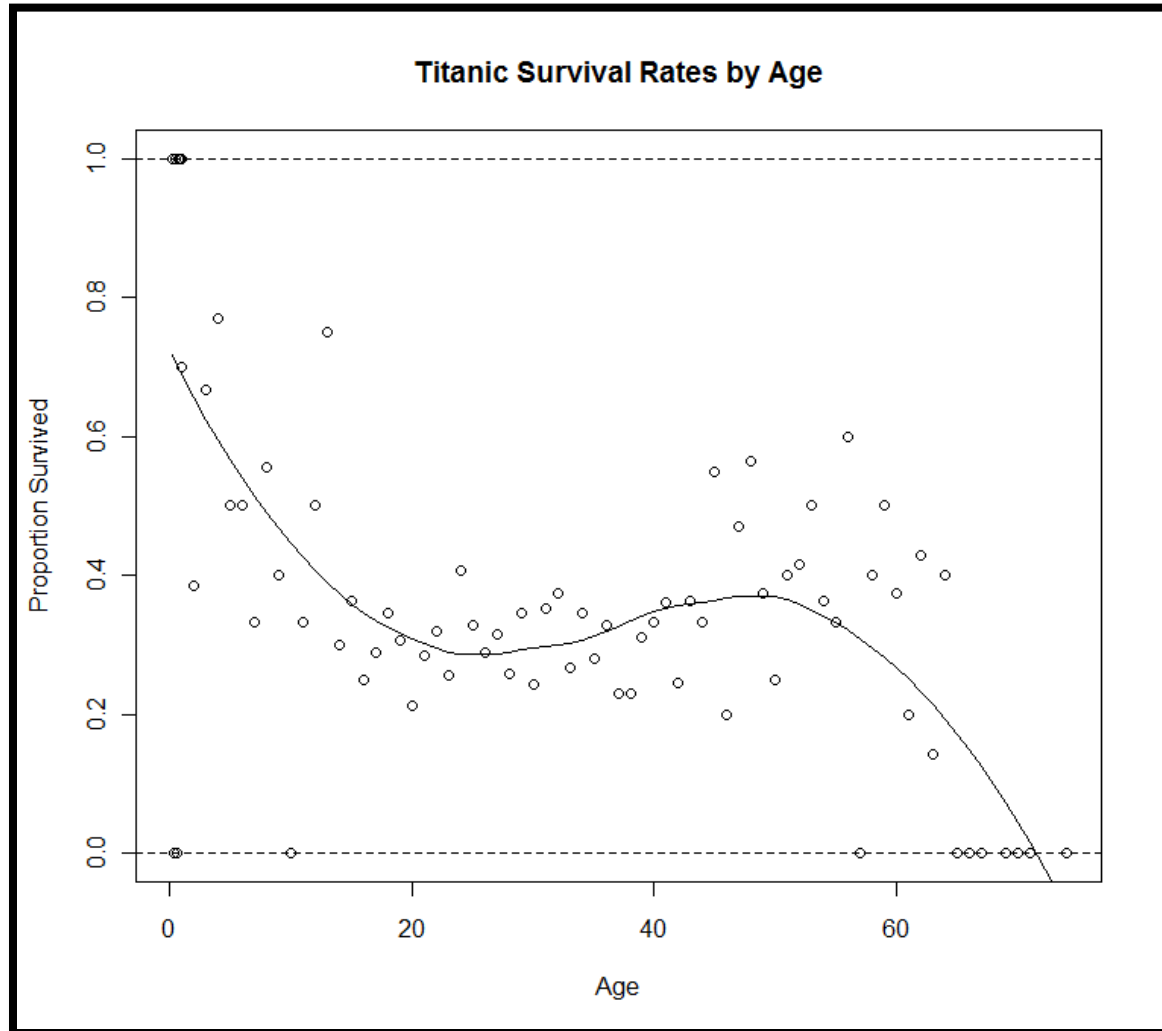
Design Improvements

I improved the visualization in the redesigned graph. Make the graph look fuller and make it easier for readers to understand the role of this graph. Secondly, I added two charts in the Micromap. The first is a bar chart, which is a graph with a percentage value that compares the changes in the population of each state over two years. The second is the arrow chart, which is the value of the population change of the two states as a value that can explain which state has changed in population, and whether it shows positive or negative growth from 2010 to 2017.

In general, redesigned graph can make the graphics have a big change. In this design's graphics we use Micromap, which can display multiple variable relationships in one graph. So, it is possible to redesign the current graphics. You can view the geographic location in the graph, and you can also view the population, population growth rate, population change, etc. of each state in each year.

Graph 3

Original Graph



(Figure: Original Graph based on the Titanic Dataset)

Description

The sinkage of Titanic was historic for the entire mankind. The incident horrified the people worldwide. After this incident, not only the Company who built Titanic was enquired but also the management was enquired too. Well the reason for it was the life

boats which were present in already lesser number were made available to higher Class people, i.e. the ones who were travelling in the First Class. And hence the Second Class and Third Class could not escape the ship leading to thousands of deaths. Many reporters claimed this thing up, research teams were set up to investigate the problem and one such research time found it to be true and formed a dataset based on those observations.

The dataset was available on the website '<https://www.kaggle.com/c/titanic>'

The first few rows of the dataset looks something like this:

PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
1	0	3	Braund, Mr. Owen Harris	male	22	1	0	A/5 21171	7.25		S
2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Thayer)	female	38	1	0	PC 17599	71.2833	C85	C
3	1	3	Heikkinen, Miss. Laina	female	26	0	0	STON/O2.	7.925		S
4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35	1	0	113803	53.1	C123	S
5	0	3	Allen, Mr. William Henry	male	35	0	0	373450	8.05		S
6	0	3	Moran, Mr. James	male			0	330877	8.4583		Q
7	0	1	McCarthy, Mr. Timothy J	male	54	0	0	17463	51.8625	E46	S
8	0	3	Palsson, Master. Gosta Leonard	male	2	3	1	349909	21.075		S
9	1	3	Johnson, Mrs. Oscar W (Elisabeth Vilhelmina Berg)	female	27	0	2	347742	11.1333		S
10	1	2	Nasser, Mrs. Nicholas (Adele Achem)	female	14	1	0	237736	30.0708		C

(Figure: Titanic Sample Dataset)

Design Flaws:

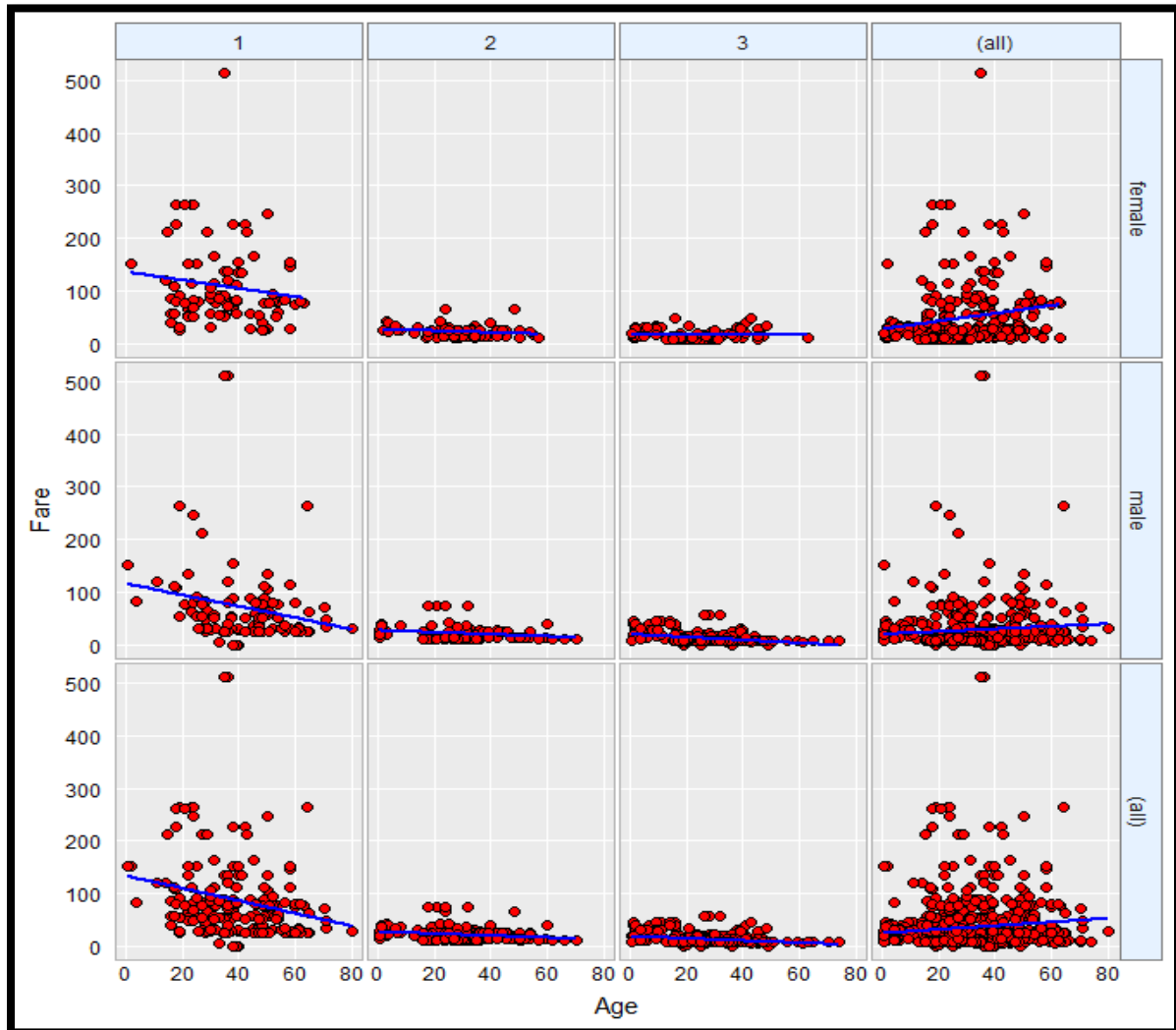
The old graph was created by Norma Enochs and was produced on the website '<http://www.statisticalconsultants.data-m.html>'

As it is visible the older graph does not have any proper specification for the Class the passengers were travelling. Also, it does not determine the Sex of the passenger travelling and which survived. The scaling's are also not proper on the graph and hence it was improper to consider it.

The graph has very poor demographics with a clumsier visualization.

Redesigned Graph

The redesigned graph was redesigned in a facet grid manner with a better graphical perspective. The redesigned graph looks like :



(Figure: Titanic Redesign Graph using Facet Grid)

The newer redesigned graph has the grids in which the data is organized in a proper way. The Data is bifurcated into classes which are 1st, 2nd and 3rd. Also, the division is based on the sex. The age is on the X axis and the fare is on the Y axis and a scatter plot along with a smoothen legend is redesigned.

Here the image speaks up for itself concluding that the more the traveller paid, the higher class he was in and hence it was more likely on his part to survive it.

References:

- "Average Salary by State." Data USA. Accessed October 26, 2018. https://datausa.io/map/?level=state&key=avg_wage,avg_wage_moe&translate=599.8074966380893,560.6547478845441&scale=2580.564535360489.
- "Economy of the United States." Wikipedia. October 20, 2018. Accessed October 26, 2018. https://en.wikipedia.org/wiki/Economy_of_the_United_States.
- Titanic | Kaggle. Accessed October 26, 2018. <http://www.kaggle.com/titanic>.
- Data Access and Dissemination Systems (DADS). "Your Geography Selections." American FactFinder - Results. October 05, 2010. Accessed October 26, 2018. https://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?pid=PEP_2017_PEPANNRES&src=pt.