ETC5521 Assignment 1

Dea Avega Editya   Siyi Li

2020-08-28

This assignment is for ETC5521 Assignment 1 by Team Quokka comprising of Dea Avega Editya and Siyi Li.

Introduction and motivation

We are motivated to explore history of slavery in the United States of America (USA) and how this dark moment changed over time of hundred years. Through the data that sourced from tidytuesday github (https://github.com/rfordatascience/tidytuesday/blob/master/data/2020/2020-06-16/readme.md), we will look at how the situation at the moment was and what can we learn from the history that maybe has correlation with current racist behavior toward black people especially in the USA.

This report has several limitations:
1. Several datasets used are only approximation of total number, i.e. route of slaves transportation is based on approximation inferred from all trans-Atlantic voyages records, and slaves names data only covers numbers of slaves who had been saved during expedition. Hence, it may not capture real situation during slavery history.
2. Census data only capture US demographics from 1790 to 1870 which is quite short regarding long existence of slavery before the census period. In addition, West region only has census data started from 1850.
3. Some proportion of data has N/A value and errors (i.e. some slaves' heights are recorded as 0 inch) which maybe ignored during data exploration.

Data description

Our datasets are retrieved from github repository of **tidytuesday** project, which has original source from US Census's Archives, Slave Voyages, and Black Past.

The data consists of 4 datasets as follows:

1. Census (in csv format) The dataset record the total slave populations across the USA during the slavery era and has 8 variables (region, division, year, total, white, black, black_free and black_slave) and 102 observation. The data is collected from a historical US census data with time period from 1790 to 1870.

2. Slave_routes (in csv format) The dataset has 8 variables (voyage_id, ship_name, port_origin, place_of_purchase, port_arrival, year_arrival, n_slaves_arrived, and captains_name) and 31.660 observations. The data is collected by approximating the total number of voyages carrying slaves from Africa, inferred from the 36,000 trans-Atlantic voyages logs. It covers information about transportation of slaves from 1514 to 1866.

3. African_names (in csv format) The dataset has 11 variables (id, voyage id , name, gender, age, height, ship name, year arrival, port embark, and country origin) and 91.490 observations. The data is collected from liberated slaves by recording their names and ages. The record is from 1808 to 1862.

4. Blackpast (in csv format) The dataset covers details around some events related to African-Americas history during slavery era until post-slavery including violence and racism events and celebrations of achievements. It has 6 variables (year, event, subject, country, state and era) and 896 observations. The data is compiled by blackpast organization (blackpast.org) from 1492 to 2009.

The wrangling process will be conducted by combining datasets of african_names and slave_routes by using "voyage_id" as key variable. In addition, this report will track slavery population in the USA based on census dataset, including average proportion of white and black slavery as well as proportion of free black and black slaves in US regions. The blackpast dataset can be used to add some detail information about particular events that maybe found interesting during data exploration.

Therefore, using these mentioned datasets this report may find explanations on a main question and three secondary questions. The primary question is **does the long slavery history in the USA explain current racism trend in the country**?

This primary question can be found after answering several secondary questions which are: 1. What is the demographic of slaves? 2. Which regions of the USA that had mostly exploited the practice? 3. Which regions of the USA that most challenging for african-americans?
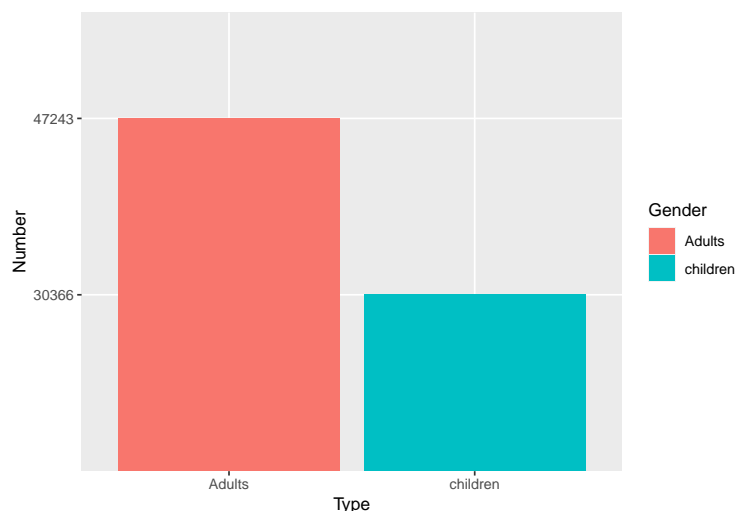
References: 1. Tidytuesday (https://github.com/rfordatascience/tidytuesday/blob/master/data/2020/2020-06-16/readme.md), 2. Slaves voyage (https://slavevoyages.org/) 3. Blackpast (https://www.blackpast.org/african-american-history-timeline/) 3. US census data (https://www.census.gov/content/dam/Census/library/working-papers/2002/demo/POP-twps0056.pdf)
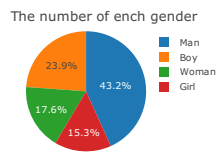
Analysis and findings
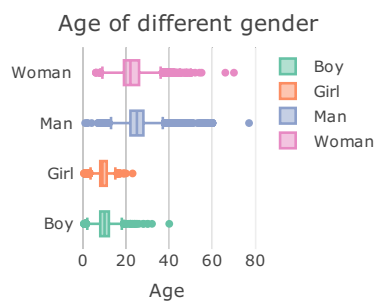
Demographic of Black Slaves

According to the gender we can divide them into 2 different part, and they are Adult and Children. There are 30,366 children while the number of the adults are 47,243.The difference between them around 17,000. We can get more information from the pie chart. Man occupied the largest percentage among the gender varible. On the contrary, the Girl has the smallest percentage.

When we look at the age of the different gender, it is easy to find that the some people just aged 0.5. It is pretty young. One posible reason of that phenomenon is that their parent is one of the slaves. Majority of the people are between **15 years old and 35 years old**. This is because people in this age tend to more productive, as they were brought to be workers in the country.
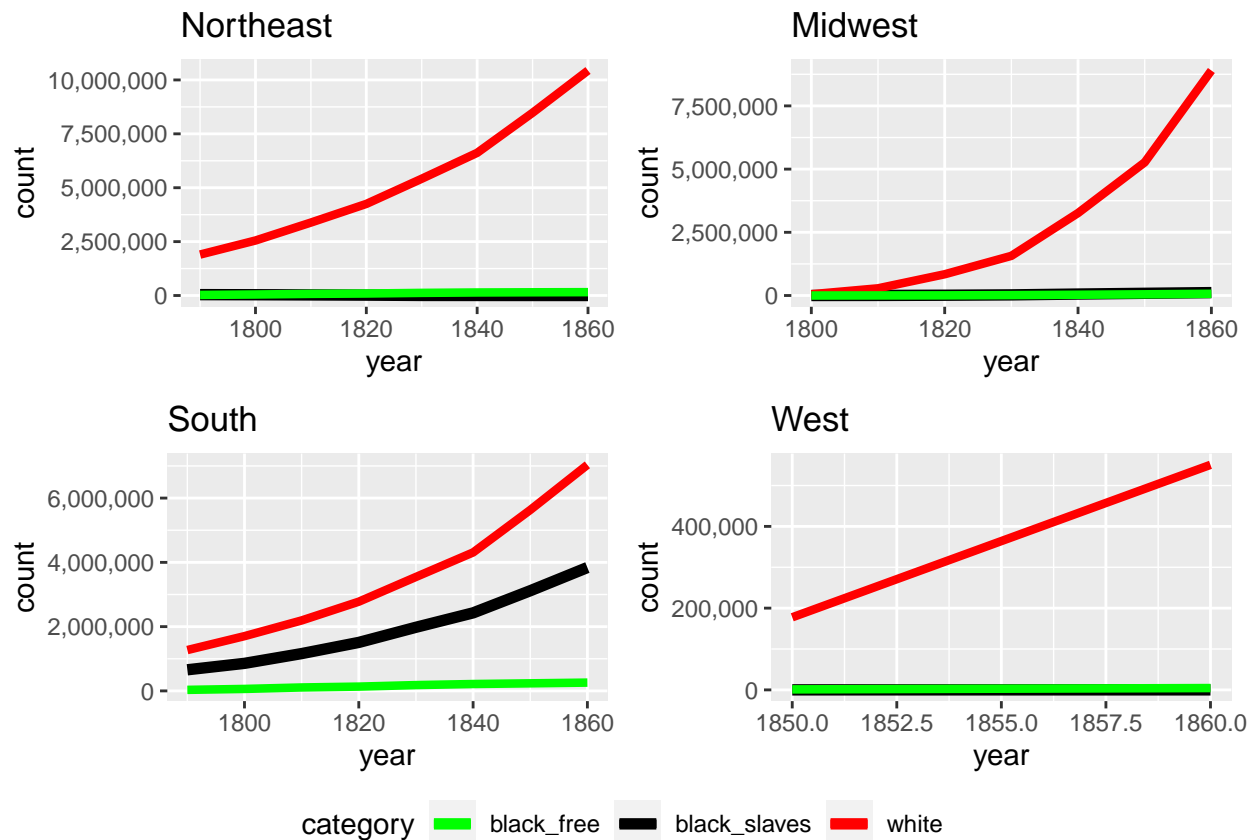
```
## Warning: 'arrange_()' is deprecated as of dplyr 0.7.0.
## Please use 'arrange()' instead.
## See vignette('programming') for more help
## This warning is displayed once every 8 hours.
## Call 'lifecycle::last_warnings()' to see where this warning was generated.
```

The number of ench gender
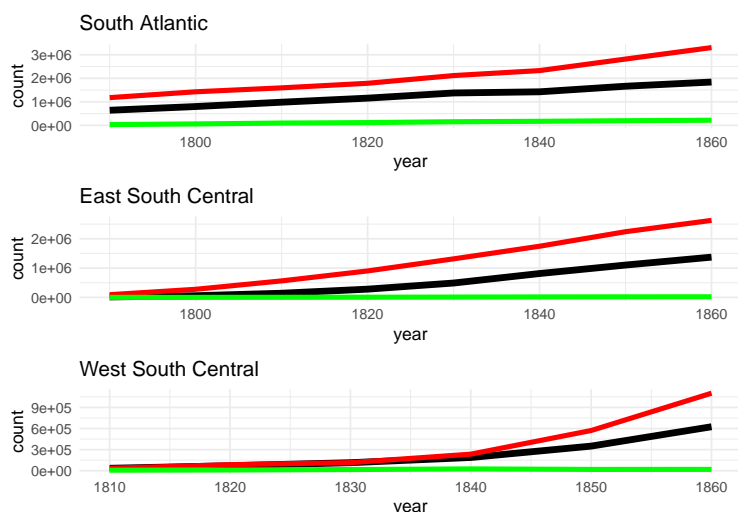
Age of different gender
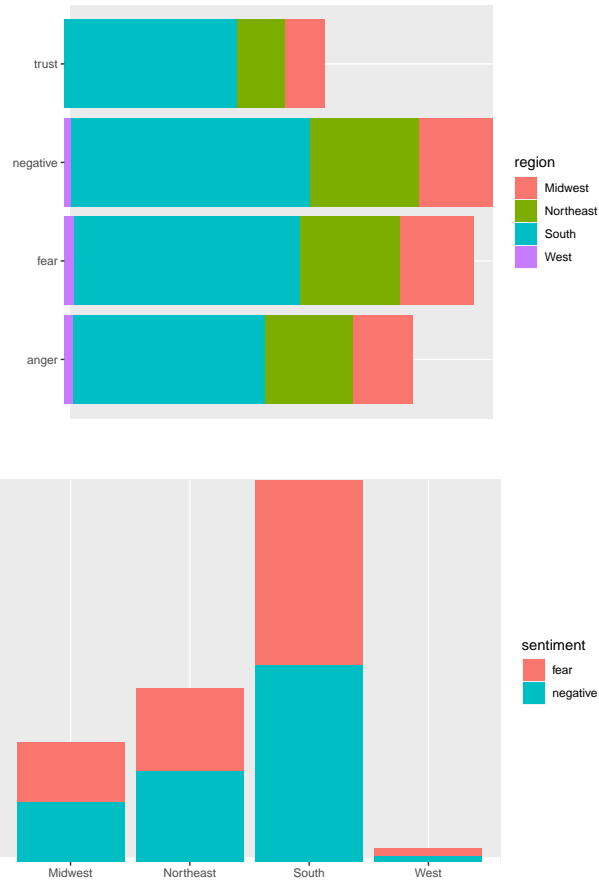
The most exploiting region in the USA

For answering this question, we will not look at the number of black slaves in each region. Instead, we will see the trend composition of these three categories, as seen in figure @ref(figure:compare-region).

According to the plot, South region has the most contrast pattern which distinguish the region from others. The population of white is increasing from 1800 up to 1860 just like what occurs in other regions. The different here is that black slaves also grows following the rise in population of white people.

The most challenging region for African-Americans

In this section, we will look at some unfortunate events that was happened in the pas using data from blakcpast dataset. For the purpose, we will filter country of interest to be only the USA.

Using these method, we can extract several high frequent bad events for black people in USA. According to the analysis in figure @ref(figure:sentiment_region), most nuance of events are **negative, fear and anger**. Furthermore, most events recorded were occured in South region. It is not surprising given the previous finding that South region is a region that seems really exploiting slavery culture.

We then focus on particular bad events which represented by nuance of disgust, fear, negative and sadness. Using these sentiments, we can see and confirm that South and Northeast region contributed to unfortunate events for black people. On the other hand, **West region** maybe a good place to live for them, due to very few bad incidents happened there.

Summary

The history of slavery has deeply rooted in some parts of the USA, particularly in South region. The slaves were brought into the country mainly for being workers, therefore the majority of slaves are **15 years old and 35 years old**

According to sentiment analysis from blackpast record of events, we can see that most bad events took place in South region followed by Northeast region. This finding coincidentally match with a 2015 research of Chae et al. (2015) that describes these two regions as the most racist regions in the United States, by calculating

number of google searches in the country that contain the N-word. (https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0122963#pone.0122963.ref001)

Bibliography

Chae, David H., Sean Clouston, Mark L. Hatzenbuehler, Michael R. Kramer, Hannah L. F. Cooper, Sacoby M. Wilson, Seth I. Stephens-Davidowitz, Robert S. Gold, and Bruce G. Link. 2015. "Association between an Internet-Based Measure of Area Racism and Black Mortality." Edited by Hajo Zeeb. *PLOS ONE* 10 (4): e0122963. https://doi.org/10.1371/journal.pone.0122963.