ETC5521 Assignment 1

Dea Avega Editya   Siyi Li

2020-08-27

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 dark history of slavery in United States of America and how this 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. 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 United States 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.
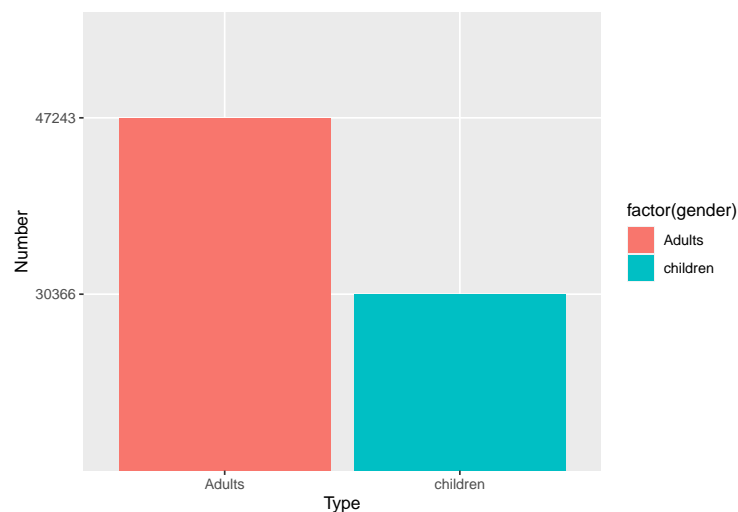
Therefore, using these mentioned datasets this report may find explanations on a main question and three secondary questions. The primary question is **does long slavery history in the USA explain current racism in the country**? This primary question can be answered after answering several secondary questions which are: 1. What is the demographic of slaves? 2. Which region of the USA that most exploit the practice? 3. How unfortunate is for being a black in the USA?

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)
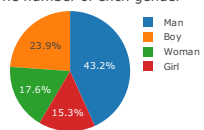
Analysis and findings

[FILL] Should include at least one plot or numerical summary for each of your questions, that helps the reader arrive at an answer. You should also write paragraphs describing the methods, summaries and findings.
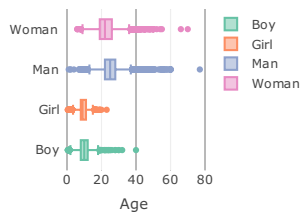
Slaves Demography


```
## 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.
```
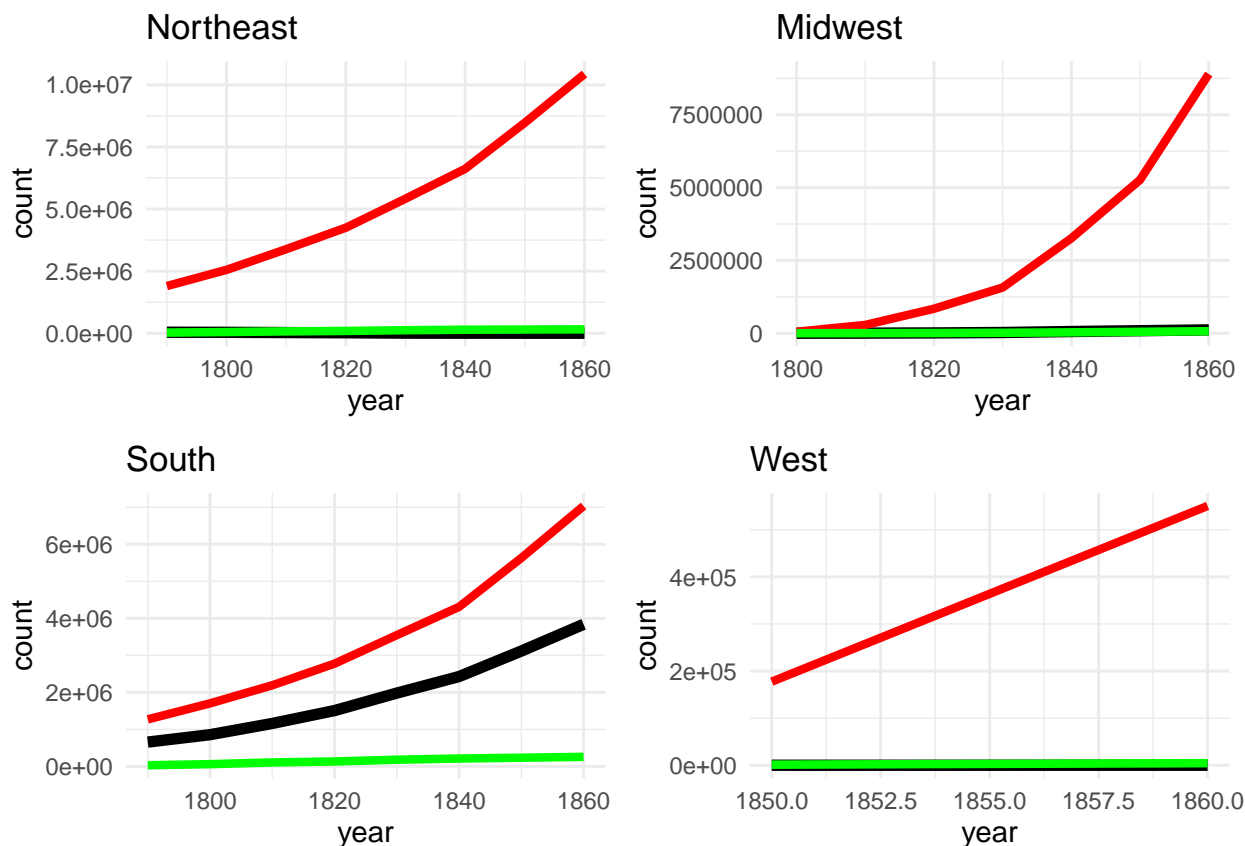
The number of ench gender

- Man
- Boy
- Woman
- Girl

43.2%
23.9%
17.6%
15.3%

Age of different gender

- Boy
- Girl
- Man
- Woman

Age

Which region/state in the USA that most exploit the slavery practice?

Seems like South region has staggering pattern which distinguish the region from the others. As we can see from figur @(ref:compare-region)

Interestingly, this past history of slavery seems deeply rooted hence can have a line with current time. According to a research paper, the most racism region in United States is in the South region based on searching according to a journal (https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0122963#pone.0122963.ref001) which calculate google searching by americans that containing the N-word.

How unfortunate is for being a black in the United States of America?

In this section, we will look at some unfortunate events that was happenned 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 unfortunate events for black people in USA. According to the analysis in figure @ref(figure:sentiment_region), most events are **negative, full of fear and anger**. Furthermore, most events recorded were occured in South region. It is not surprising given previous finding that South region is the most slavery-exploiting region.

Looking at figure @ref(figure:sentiment_negative), we can see more clearly the regions where bad events mostly took place. As shown in the chart, the winner is South region followed by the second "winner" which is the Northeast region. A good place to live for black people perhaps is West region with quite few bad events recorded.

References