

# ETC5521 Assignment 1

## Unboxing the Dark History of US Slavery

Dea Avega Editya      Siyi Li

2020-08-28

## Contents

<b>1</b>	<b>Introduction and motivation</b>	<b>1</b>
<b>2</b>	<b>Data description</b>	<b>1</b>
<b>3</b>	<b>Analysis and findings</b>	<b>2</b>
3.1	Demographic of Black Slaves . . . . .	2
3.2	Finding the most exploiting region . . . . .	4
3.3	The most unfriendly region for African-Americans . . . . .	4
3.4	Summary . . . . .	6
	<b>Bibliography</b>	<b>6</b>

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

## 1 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 tidyuesday 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. Data sets used only contain relatively small observations, i.e. slaves\_name record only covers numbers of slaves saved during their expedition. Hence, it may not really capture situation during slavery history.
2. Census data only capture US demographics from 1790 to 1870 which is quite short regarding long existence of slavery (prior to the census period). In addition, West region only has census data of year 1850 and 1860.
3. Some proportion of data has N/A value and errors which would be omitted during data exploration.

## 2 Data description

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

There are four data sets in the tidyuesday's repo, however for this report's purpose we only use three data sets which are:

1. Census (in csv format) The data set 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. African\_names (in csv format) The data set 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.
3. Blackpast (in csv format) The data set covers details around some events related to African-Americans 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 is conducted by grouping some variables in the african\_names to have aggregate number for each category, hence enable us to compare across categories. We also recalculate number of total population in census data set since the existing number is somehow incorrect for West region (we find this miscalculation after visualizing the data). These proportions will be used to track slavery exploitation in the USA.

Furthermore, this report also takes advantage of quite comprehensive record of African-Americans' historical events in the blackpast data set, to analyze which region of USA that seems to be unfriendly to the African-American people related to some unfortunate events recorded.

Therefore, using all of these mentioned datasets this report will find a brief explanation on a main question: **Does the long slavery history in the USA explain current racism towards African-American?**

In order to answer the main question, we will first look at these secondary questions: 1. What is the demographic of black slaves? 2. Which region of the USA that had most exploited the practice? 3. Which region is unfriendly to black people?

References of data sets sources:

1. TidyTuesday (<https://github.com/rfordatascience/tidytuesday/blob/master/data/2020/2020-06-16/readme.md>),
2. 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>)

## 3 Analysis and findings

### 3.1 Demographic of Black Slaves

```
## 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.
```

In this section, we will extract some information from african\_names dataset related to demographic of slaves. By extracting and analyzing the data, we can have initial picture of enslavement practice before moving on to further analysis in the following sections.

We group the category of slaves by **id** (since we are not interested to know their names and sometimes people can have similar names), **gender** and **age**. Thus, we can get information about gender category from a pie chart of figure 2. The plot shows us that men occupy the largest percentage of total slaves in the observed data, followed by boys. Proportion of boys is also larger than women. Meanwhile, girls contribute to the smallest percentage of slaves.

In general, we can see that men and boys' proportion are significantly larger than female (women and girls). However, we realize that if we look at children proportion (boys and girls) as shown in figure 1, it is really sad to know that they make up almost a half of the total slaves.

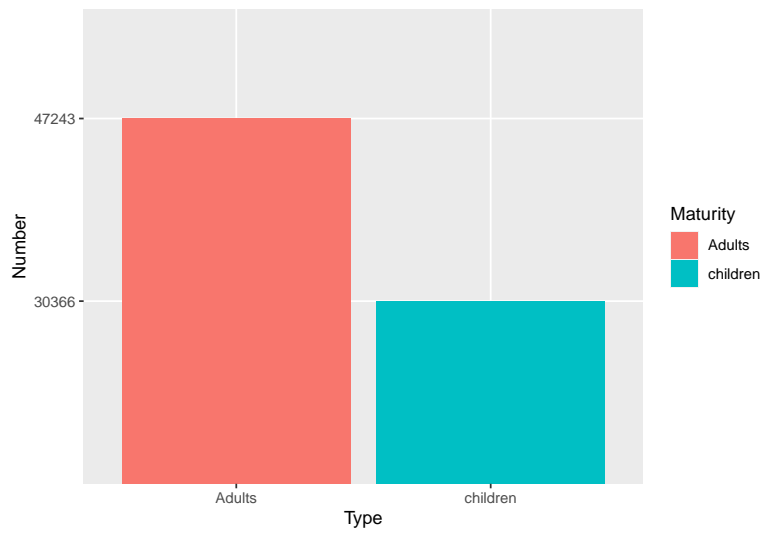


Figure 1: Maturity of Slaves

### Composition of Slaves Based on Gender

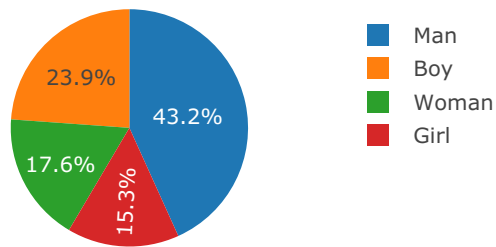


Figure 2: Composition Based on Gender

### Age of different gender

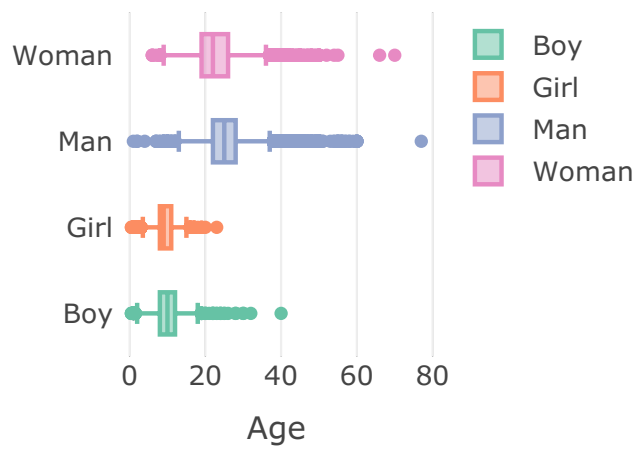


Figure 3: Age Distribution

When we look at the age distribution of slaves as shown in figure 3, majority of the slaves are between **15 and 35 years old**. This range of productive ages is not surprising since they were brought into the country mainly to be a worker, according to prior information we got from blackpast.org.

Though, there are some outliers that can be spotted from the figure 3 which we suspect due to errors in data recording, i.e. a boy has age of 40 and a man is 77 years old (which we think way too old to be a slave). Another finding is there are some slaves that just age 5 months old as well as some children under 5 years old, which are absolutely pretty young. One possible reason is that their parents are among the slaves that brought along their children.

### 3.2 Finding the most exploiting region

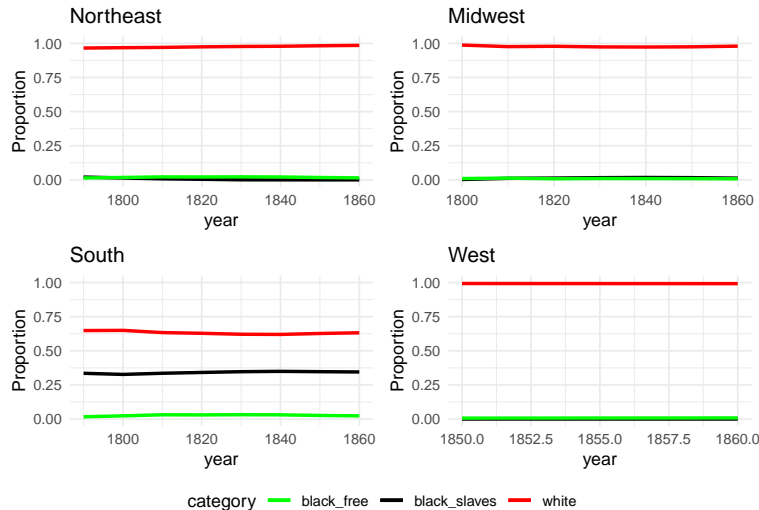


Figure 4: Composition of White and Black

For answering the question of which region that most exploits the slavery practice, we will not look at the number of black slaves in each region of the USA. Instead, we will see the trend of proportion from these three categories (white, black slaves, and black free), as seen in figure 4. In the figure, We are not covering census year of 1870 since the enslavement ended in that year.

Seeing the plot, we can easily spot that **South region** has the most contrast pattern which distinguish the region from the others. The proportion of white people in South region slightly decrease from 1800 until 1860, however the proportion of black slaves is growing and comprises more than a quarter of the total population in that region.

After spotting the pattern, we are interested to explore further in the South region to see the slavery practice in division level. South region consists of three divisions, South Atlantic, East South Central and West South Central. From figure 5 we can see clearly that East South Central division has more progressive pattern of slavery exploitation during the observed period, where other divisions tend to have more stable pattern of slavery practice. Nevertheless, all of these divisions have quite similar proportion of black slaves in the last census year of slavery period (1860).

### 3.3 The most unfriendly region for African-Americans

In this section, we will connect the past and post-slavery period in the USA through some observed events from blackpast data set before answering our main question. For the purpose, we will first filter country of interest in the data set to be only the USA. Later on, we will select only important words (filtered using English stop words of tidytext R package (Silge and Robinson (2016))). Having all essential words, we will try to find various sentiments in these words by using NRC sentiment (Mohammad (2018)) and grouping the sentiments by region.

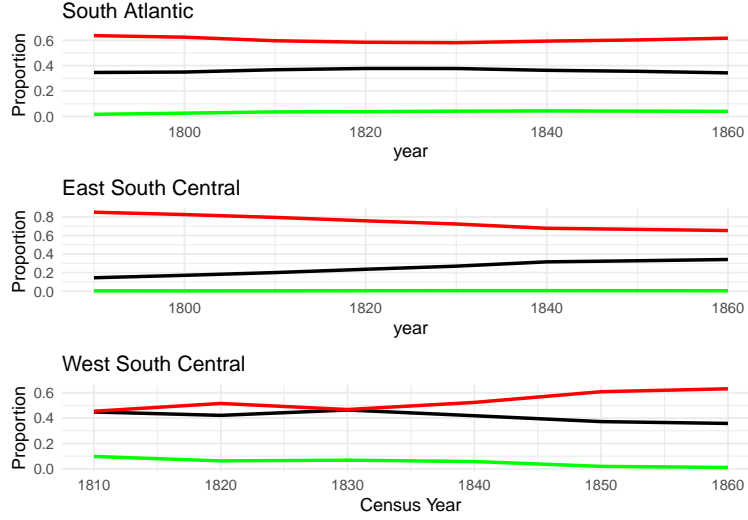


Figure 5: Proportion of White and Black

In order to link this analysis with the previous section, we add variable of regions that correspond to the state names in the blackpast data set. For example, we add Northeast for covering states like Connecticut, New York and New Jersey. In addition we also focus on all events that related to slavery and racism behavior toward African-American, represented by subjects such as “**Slave Laws**”, “**Slave Labor**”, “**Racial Restrictions**”, “**Racial Violence**”, “**Resistance to Enslavement**”, “**The Slavery Controversy**”, and “**Antebellum Slavery**”).

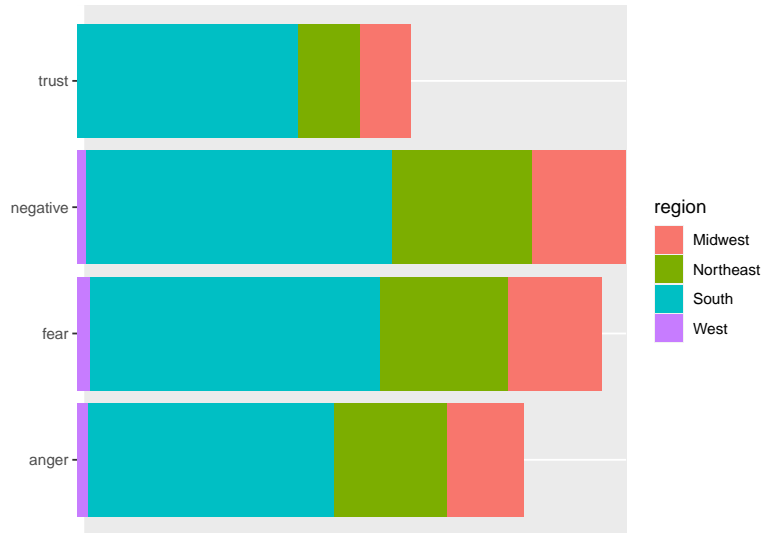


Figure 6: Sentiment Analysis in Regions

Using these method, we can observe occurrence of bad events in the USA history which mostly comprises of racial restrictions and racial violence. According to the analysis in figure 6, most nuance of observed events are related to **negative, fear and anger** feelings. Moreover, most events recorded were occurred in the South region. Again, it is not surprising given the previous finding that South region is the region that seems really exploiting slavery practice.

We then more focus on particular bad events which represented by sentiment of “disgust”, “fear”, “negative” and “sadness” (we get rid of positive sentiments such as joy) and plot it into bar chart as seen in figure 7.

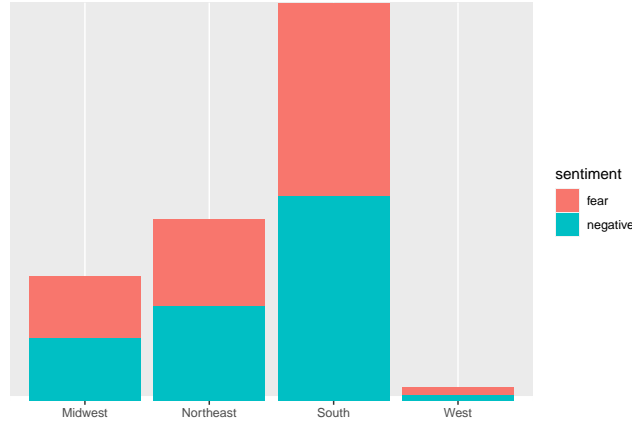


Figure 7: Negative Nuances

Using these negative sentiments, we can see that South and Northeast region are seems like “unfriendly” to the African-American people, since both regions contribute to most of these bad events in the observed data. On the other hand, **West region** maybe a good place to live for African-American, due to very few bad incidents happened there.

### 3.4 Summary

The history of slavery has deeply rooted in some parts of the USA, and the practice was mostly exploited in **the South region** following our data exploration in the second section. The slaves were brought into the country mainly for being a worker, therefore most slaves were in their productive ages, ranges from 15 to 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**. These bad events recorded are related to slavery exploitation and racism behavior/racial violence (in post-slavery era) towards Africa-American. Hence, we may argue that today’s racism possibly has association with long history of enslavement, especially in the case of South region.

This finding is quite intriguing since it is inadvertently backed by a journal research from Chae et al. (2015) that describe these two regions (South and Northeast) as the most racist regions in the United States.

## Bibliography

- Arnold, Jeffrey B. 2019. *Ggthemes: Extra Themes, Scales and Geoms for 'Ggplot2'*. <https://CRAN.R-project.org/package=ggthemes>.
- Auguie, Baptiste. 2017. *GridExtra: Miscellaneous Functions for "Grid" Graphics*. <https://CRAN.R-project.org/package=gridExtra>.
- 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>.
- Firke, Sam. 2020. *Janitor: Simple Tools for Examining and Cleaning Dirty Data*. <https://CRAN.R-project.org/package=janitor>.
- Grolemund, Garrett, and Hadley Wickham. 2011. “Dates and Times Made Easy with lubridate.” *Journal of Statistical Software* 40 (3): 1–25. <http://www.jstatsoft.org/v40/i03/>.

- Kassambara, Alboukadel. 2020. *Ggpubr: 'Ggplot2' Based Publication Ready Plots*. <https://CRAN.R-project.org/package=ggpubr>.
- Mohammad, Saif M. 2018. *Word Affect Intensities*. Miyazaki, Japan.
- Sievert, Carson. 2020. *Interactive Web-Based Data Visualization with R, Plotly, and Shiny*. Chapman; Hall/CRC. <https://plotly-r.com>.
- Silge, Julia, and David Robinson. 2016. “Tidyttext: Text Mining and Analysis Using Tidy Data Principles in R.” *JOSS* 1 (3). <https://doi.org/10.21105/joss.00037>.
- Tierney, Nicholas, Di Cook, Miles McBain, and Colin Fay. 2020. *Naniar: Data Structures, Summaries, and Visualisations for Missing Data*. <https://CRAN.R-project.org/package=naniar>.
- Wickham, Hadley. 2007. “Reshaping Data with the reshape Package.” *Journal of Statistical Software* 21 (12): 1–20. <http://www.jstatsoft.org/v21/i12/>.
- Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D’Agostino McGowan, Romain François, Garrett Golemund, et al. 2019. “Welcome to the tidyverse.” *Journal of Open Source Software* 4 (43): 1686. <https://doi.org/10.21105/joss.01686>.
- Wickham, Hadley, Romain François, Lionel Henry, and Kirill Müller. 2020. *Dplyr: A Grammar of Data Manipulation*. <https://CRAN.R-project.org/package=dplyr>.
- Wickham, Hadley, Jim Hester, and Romain Francois. 2018. *Readr: Read Rectangular Text Data*. <https://CRAN.R-project.org/package=readr>.