

# Analyzing and recreating data visualizations by W.E.B. Du Bois

[ADD DISCLAIMER ABOUT ANTIQUATED LANGUAGE USED]

## Introduction

In February 2021, Allen Hillery, Athony Starks, and Sekou Tyler, started the #DuboisChallenge, and annual online challenge where participants use modern data visualization tools such as R, Python, and Tableau, among others, to recreate the data visualizations by W.E.B. Du Bois in the 1900 *Exposition Universelle* (also known as the 1900 Paris Exposition). The seven-week challenge included 10 out of the 63 visualizations in the original exhibit. Each week, participants were tasked with recreating one of the visualizations and there were three “bonus” visualization challenges. People used social media to share their recreations side-by-side with the originals using the tag #DuBoisChallenge, and many shared the code they used for the recreation. The challenge took place again in February 2022 and it was featured in 2021 and 2022 as part of RStudio’s weekly online data visualization event, TidyTuesday.

Many instructors in statistics and data science courses present the historical work of Florence Nightingale, a pioneer of data visualization, as early examples of visually presenting data in a compelling way. Less often presented, however, is the work of W.E.B. Du Bois, a sociologist, author, and activist, who used data visualizations to share the lived experience of African Americans in the late 19th and early 20th centuries. Though not a statistician by training, the influence of his work can still be seen in the visualizations produced by the statistics and data science communities today. As the statistics community (and STEM more broadly) strives to increase representation by people in historically marginalized groups, there is an opportunity to share and celebrate the historical work of pioneers like W.E.B. Du Bois in the classroom.

In this column we present examples of activities for the classroom that are inspired by the #DuBoisChallenge. We begin with a brief history of the 1900 Paris exhibit. Then we present two activities for introductory-level students based on the visualization “Proportion of Total Negro Children of School Age Who Are Enrolled in the Public Schools.” We walk through these activities in detail, sharing discussion questions as students interpret the graph along

with guidance to lead students through recreating the visualization. Next we present examples of two more recreation activities that can be used as a follow-up to the first or for students in more advanced-level data visualization courses. All of the activities are presented using `ggplot2` in R; however, the same concepts could be applied to recreate the visualizations using other software.

The data for the visualization and R code to recreate the visualizations presented in this column are available in the GitHub repository at [bit.ly/taking-a-chance](https://bit.ly/taking-a-chance).

## A Brief History

The book *W.E.B. Du Bois's Data Portraits: Visualizing Black America* edited by Whitney Battle-Baptiste and Britt Rusert, gives a history of Du Bois's exhibit at the *Exposition Universelle* (also known as the 1900 Paris Exposition) along with prints and details of the 63 visualization presented at the event. We provide a brief summary of the historical background in the book that instructors can share with students as background and context for the collection of visualizations. We encourage the readers to see this book and other sources in the *Further Reading* section for a more detailed history of Du Bois and the 1900 Paris Exposition.

Du Bois's visualizations were part of the Exposition des Nègres d'Amérique (The Exhibit of American Negroes), an exhibit organized by newspaper editor Thomas Junius Calloway. The goal of the exhibit was to tell the story of African Americans after Emancipation using a variety of items including texts, portraits, and data visualizations. The exhibit highlighted the progress made by African Americans, and Du Bois in particular, used data and visualizations to counter the “narrative of Black inferiority.”

As a professor at Atlanta University, a historically Black university in Georgia, Du Bois was one of the first professors in the United States to train students in empirical sociological involving data collection and analysis. Therefore, he worked with his current and former students to create the visualizations for the 1900 exhibit. They used data collected by Du Bois's sociology lab, government reports, and data from the United States Census. The data were used to create two sets of visualizations: one focusing specifically on the experience of African Americans in Georgia and one focusing on more national-level statistics and trends. The visualizations were hand drawn using ink, watercolor, and graphite. They stood out from other visualizations of the time with their bright colors and modern style, an intentional design choice by Du Bois to make more effectively convey the message to the Parisian audience.

Du Bois's visualizations and photographs from the 1900 Paris Exhibit are available in the [Library of Congress digital collection](#).

## Activity 1

## Activity 2

### Extension

Another example of a visualization that can be recreated using segmented bar charts is “Income and Expenditure of 150 Negro Families in Atlanta, G.A., U.S.A.” (Plate 31 in the collection). The visualization was originally created using ink and watercolors. It also stands out due to the use of photographs at the top of the graph. The visualization shows family budgets split by income classes for 150 families in Georgia. The copy of the original graph is available in [Library of Congress digital collection](#), and a recreation of the graph using `ggplot2` is below.

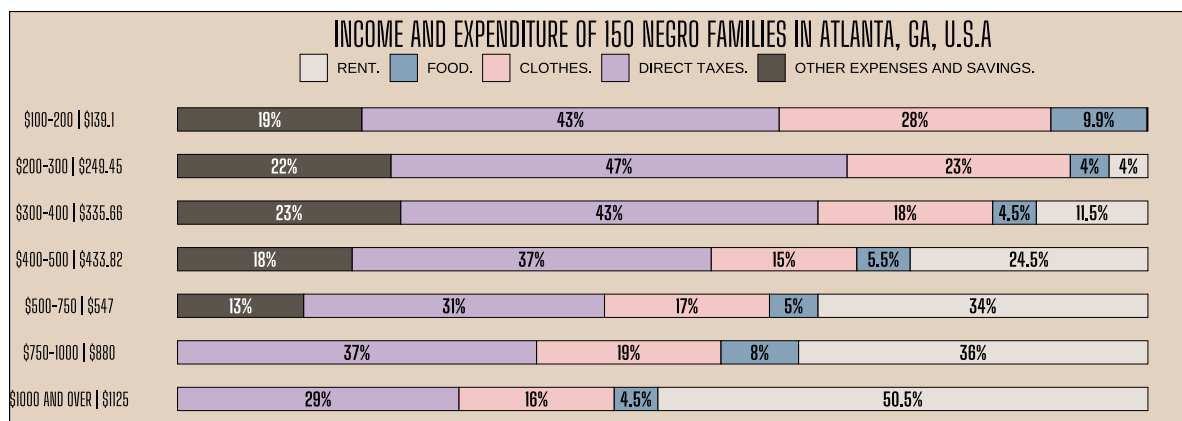


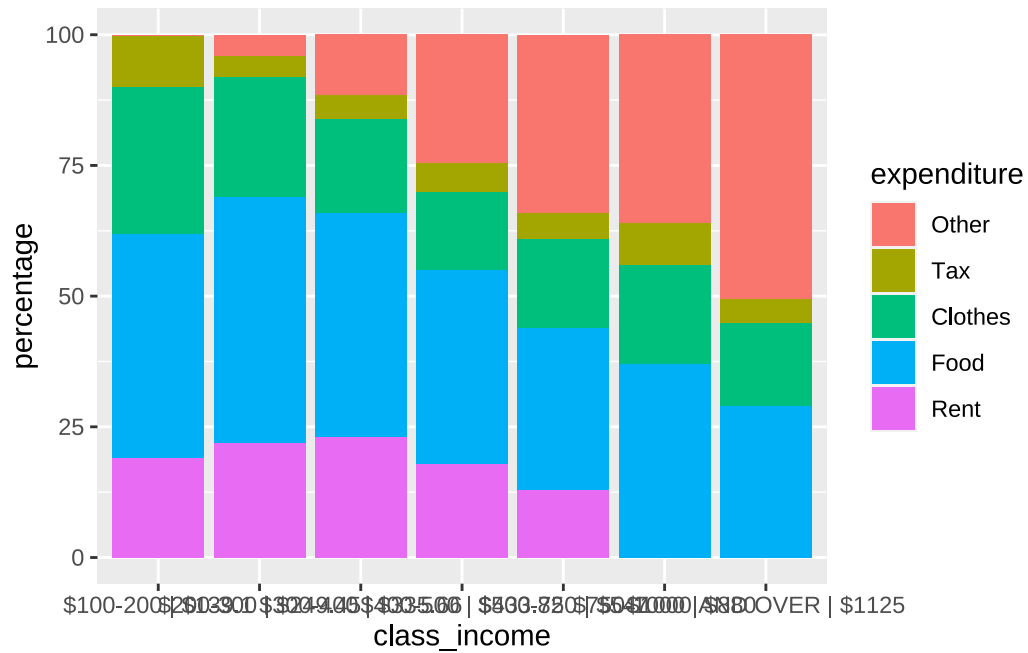
Figure 1: Recreation of “Income and Expenditure of 150 Negro Families in Atlanta, G.A., U.S.A.” (Plate 31)

We can begin by having students analyze and make observations from the original graph. Below are some discussion questions that can be used in a written assignment or to promote class discussion.

- Q1.** What is the topic of the graph?
- Q2.** What do you observe from the graph?
- Q3.** What questions do you have after viewing this graph? What do you want to learn more about?

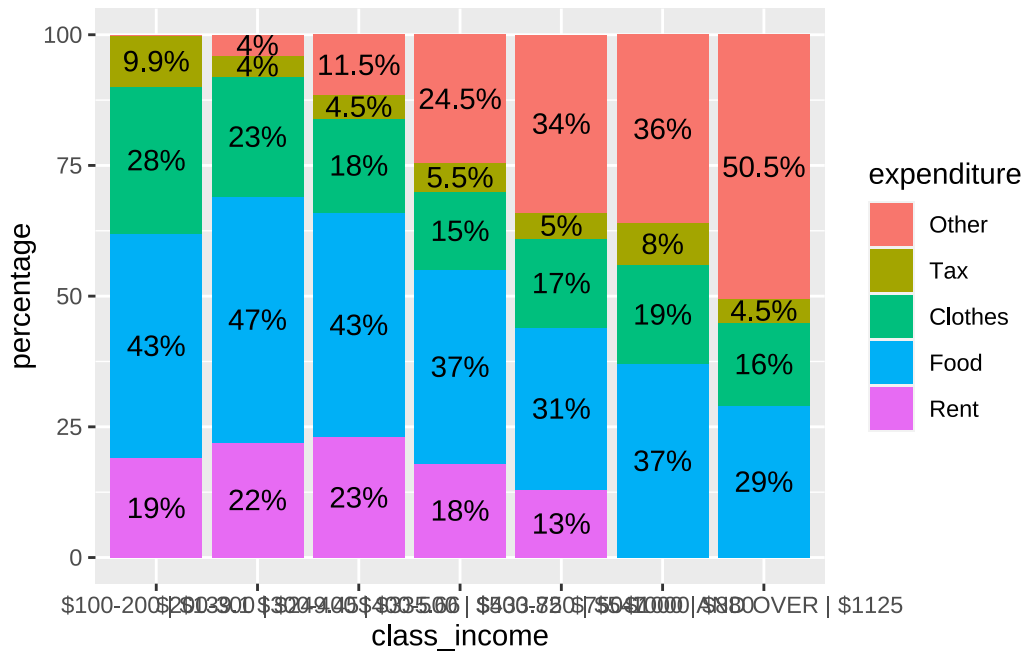
### Step 1. Create segmented bar plot

```
ggplot(data = income_long, aes(x = class_income, y = percentage, fill = expenditure)) +  
  geom_col(position = "stack")
```



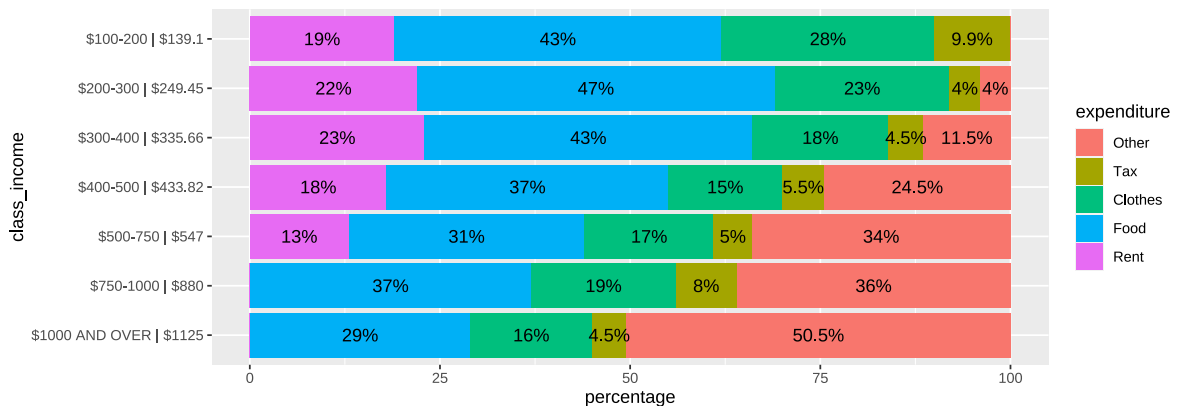
**Step 2. Add labels within each bar.**

```
ggplot(data = income_long, aes(x = class_income, y = percentage, fill = expenditure)) +
  geom_col(position = "stack") +
  geom_text(aes(label = percent_label), position = position_stack(0.5))
```



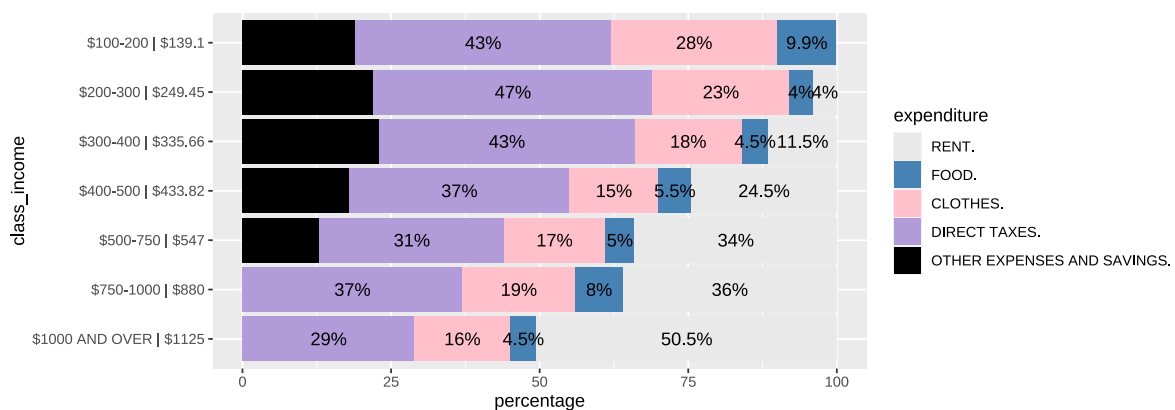
**Step 3. Flip coordinates to make bars horizontal.**

```
ggplot(data = income_long, aes(x = class_income, y = percentage, fill = expenditure)) +
  geom_col(position = "stack") +
  geom_text(aes(label = percent_label), position = position_stack(0.5)) +
  coord_flip() +
  scale_x_discrete(limits = rev)
```



#### Step 4. Use Du Bois color palette and update legend

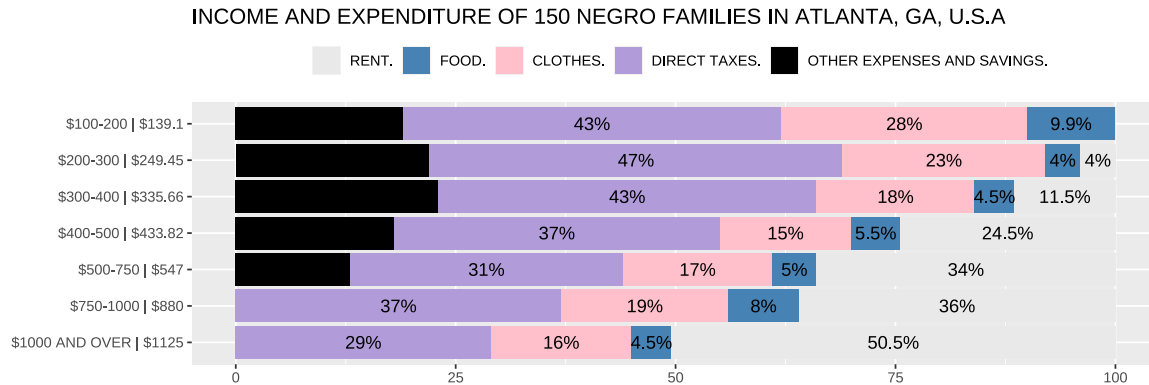
```
ggplot(data = income_long, aes(x = class_income, y = percentage, fill = expenditure)) +
  geom_col(position = "stack") +
  geom_text(aes(label = percent_label), position = position_stack(0.5)) +
  coord_flip() +
  scale_x_discrete(limits = rev) +
  scale_fill_manual(labels = c("RENT.", "FOOD.", "CLOTHES.", "DIRECT TAXES.",
                              "OTHER EXPENSES AND SAVINGS."),
                    values = c("Rent" = "#000000", "Food" = "#b19cd9",
                              "Clothes" = "#ffc0cb", "Tax" = "#4682B4",
                              "Other" = "#E9E9E9"))
```



#### Step 5. Reposition legend and update axis labels and title

```
ggplot(data = income_long, aes(x = class_income, y = percentage, fill = expenditure)) +
  geom_col(position = "stack") +
  geom_text(aes(label = percent_label), position = position_stack(0.5)) +
  coord_flip() +
  scale_x_discrete(limits = rev) +
  scale_fill_manual(labels = c("RENT.", "FOOD.", "CLOTHES.", "DIRECT TAXES.",
                              "OTHER EXPENSES AND SAVINGS."),
                    values = c("Rent" = "#000000", "Food" = "#b19cd9",
                              "Clothes" = "#ffc0cb", "Tax" = "#4682B4",
                              "Other" = "#E9E9E9")) +
  labs(x = "", y = "", fill = "",
       title = "INCOME AND EXPENDITURE OF 150 NEGRO FAMILIES IN ATLANTA, GA, U.S.A") +
```

```
theme(legend.position = "top")
```



## Additional customization

## Conclusion

## Further reading

- Battle-Baptiste, W., & Rusert, B. (Eds.). (2018). *WEB Du Bois's data portraits: Visualizing black America*. Chronicle Books.
- Recreations of Du Bois's Data Portraits GitHub repository: [github.com/ajstarks/dubois-data-portraits](https://github.com/ajstarks/dubois-data-portraits)
- Starks, A. (2019, August 21). Recreating W.E.B Du Bois's Data Portraits. Medium. [medium.com/nightingale/recreating-w-e-b-du-boiss-data-portraits-87dd36096f34](https://medium.com/nightingale/recreating-w-e-b-du-boiss-data-portraits-87dd36096f34)
- Starks, A. (2022, February 1). The #DuBois Challenge. Nightingale. [nightingaledvs.com/the-dubois-challenge](https://nightingaledvs.com/the-dubois-challenge)