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

> **!** Important
>
> The visualizations presented in this article are original data visualizations by W.E.B. Du Bois and the captions reflect the language of the time in history.

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 the weekly online data visualization event TidyTuesday (https://github.com/rfordatascience/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 Exposition. Then, we present two activities that involve the reproduction of two of Du Bois's visualizations. Each activity is presented in two parts: part 1 for an introductory audience and part 2 for a more advanced audience. 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.

We're providing these activities assuming students use R and the **ggplot2** package for data visualization. The activities can also be used for teaching with other languages or software as well as other plotting systems in R, but the syntax-specific instructions will need to be updated.

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 .

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.

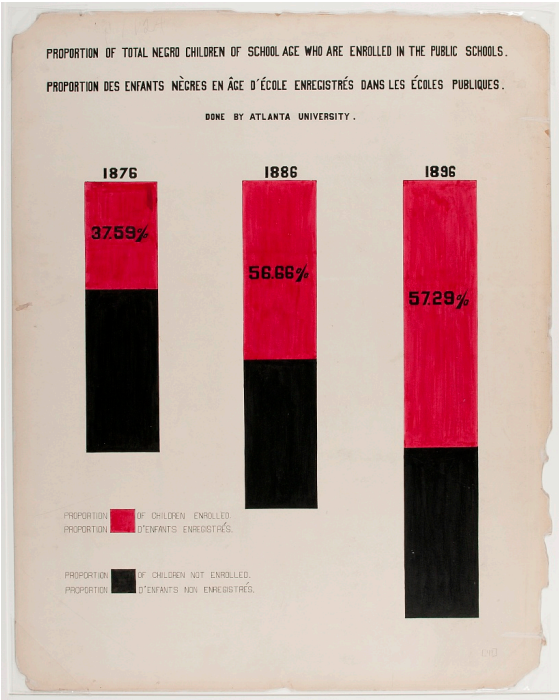
Du Bois's visualizations were part of the Exposition des Nègres d'Amérique (The Exhibit of American Negroes), an exhibit organized by newspaper editor Thomas Junius Calloway. The goal of the exhibit was to tell the story of African Americans after Emancipation using a variety of items including texts, 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.

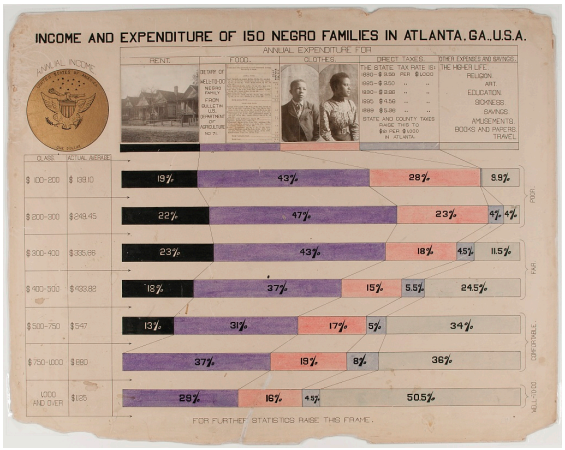
Du Bois's visualizations and photographs from the 1900 Paris Exposition are available in the Library of Congress digital collection.

Activities

The most prevalent type of visualizations created by W. E. B. Du Bois are bar charts, so the activities will focus on recreating the following, seemingly simple, bar charts.



(a) Activity 1: Proportion Of Total Negro Children Of School Age Who Are Enrolled In The Public Schools (Plate 49)



(b) Activity 2: Income and Expenditure of 150 Negro Families in Atlanta, GA., U.S.A. (Plate 31)

Figure 1: Two data visualizations by W.E.B. Du Bois.

Both activities are presented in two parts. Part 1 is aimed at the introductory statistics and data science courses that teach data visualization as part of their curriculum. In this part we focus on getting a stacked bar chart representing the correct percentages and making it look like the Du Bois visualizations, at a high level. Part 2 is aimed at a more advanced audience in a data visualization course, building on Part 1 to capture all features of the visualization, including annotations and fonts.

Activity 1 - Enrollment in Public Schools

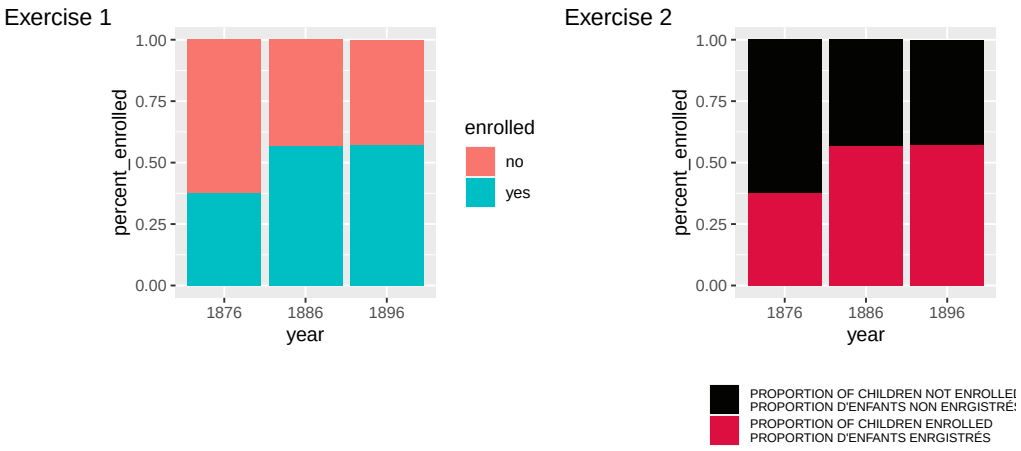
The goal of the this activity is to reproduce the visualization shown in Figure 1a, which displays the percentages of school aged Black children who are and are not enrolled public schools in the years 1876, 1886, and 1896.

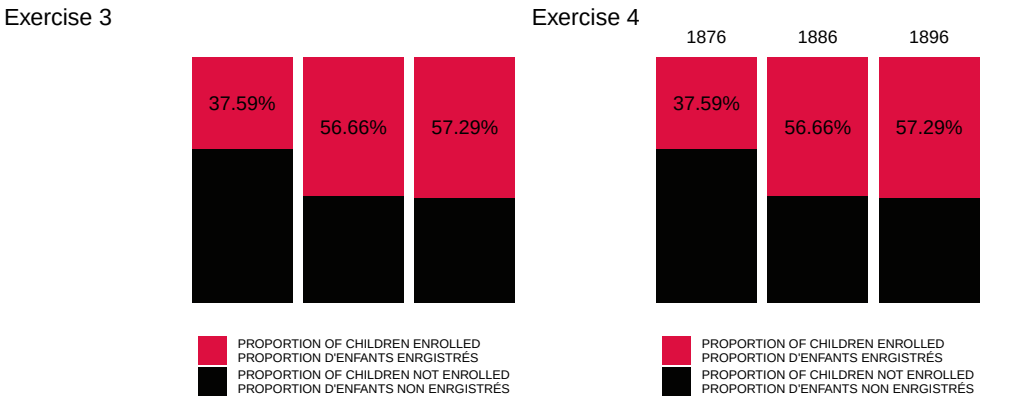
For this activity we recommend providing the following dataset to your students.

enrolled	year	percent_enrolled
yes	1876	37.59
no	1876	62.41
yes	1886	56.66
no	1886	43.34
yes	1896	57.29
no	1896	42.71

The following six exercises comprise Part 1 of this activity and they break down the task of reproducing the visualization, starting with ggplot2 defaults and customizing a little bit at each step until we reach something very similar (though not a perfect replica) of the original plot. We would recommend providing the exercises along with the target visualizations given in Figure 2.

1. Create a stacked bar chart where year is on the x-axis and percent enrolled is on the y-axis. Once you do, compare the plot you made to the original plot you're reproducing and make a list of all updates you will need to make to this plot to make it look more like for a complete reproduction.
2. Update the colors on the plot to match the inspiration figure. Also make sure that the order in which the colors show up in the bars as well as the legend labels match. Note: For this task you can either use red and black colors, or use an additional piece of software (e.g., Digital Color Meter on a Mac) to match the exact colors.
3. Place the percentages of "yes"s in the red portions of the bars and remove all axis elements by setting `theme_void()`. Hint: First calculate where on the y-axis the annotation should be placed and store those values in the data as a new column.
4. Add back the x-axis labels, but place them on top of the bars.
5. Match the plot title and subtitle.
6. Make the bars skinnier and increase the white space between them.





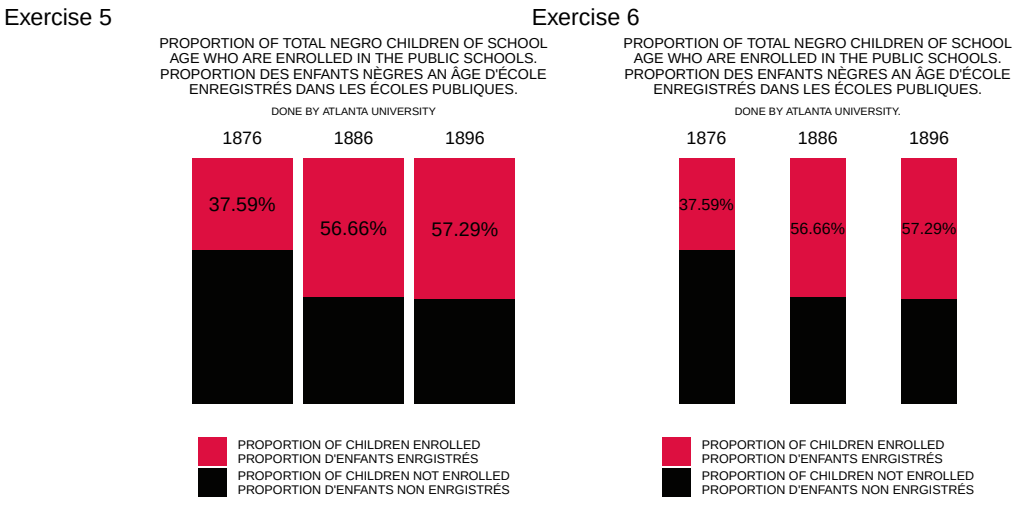
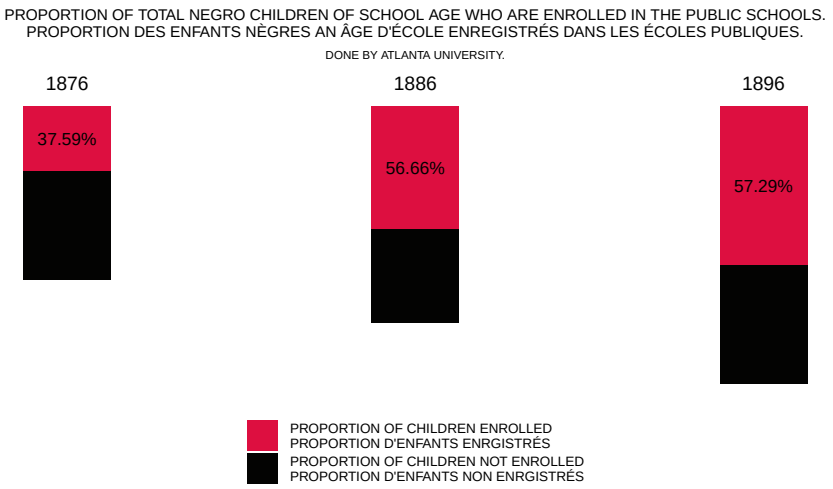


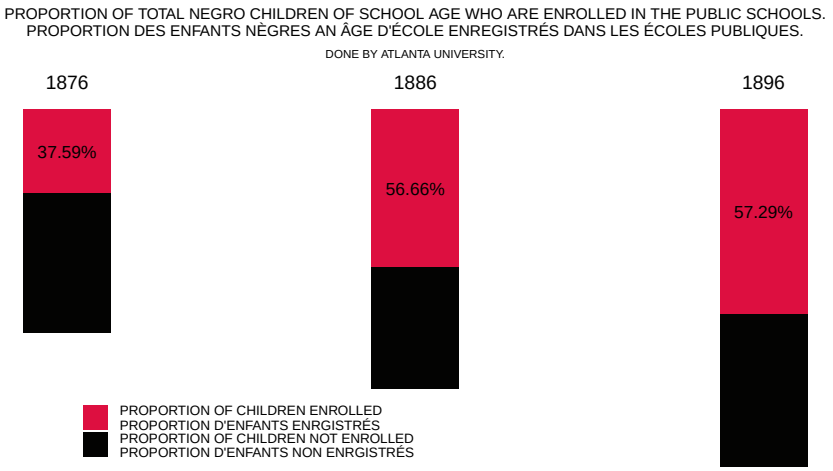
Figure 2: Figures to recreate for the exercises in Activity 1 - Part 1.

For Part 2 of this activity we'll focus on four features of the original plot from Figure 1a that we didn't replicate in Part 1: unequal bar heights, legend placement, background color and texture, and fonts. The following two exercises address the first two of these. As with the exercises in Part 1, we recommend providing these exercises along with target figures shown in

6. The original visualization in Figure 1a features three bars of unequal height. Presumably, this might indicate a larger population being represented, but we don't have data on how much larger this population is. But we can eyeball that the bar for 1886 is roughly 25% taller than the bar for 1876 and the bar for 1896 is roughly 60% taller. Recreate the visualization with unequal bar heights using these rough estimates.



7. Move the legend underneath the first bar and reduce the size of the legend key.



These modifications bring us very close to the original figure, but there are still a couple more features to try to match. The following two exercises address these.

8. Change the background of the plot to match the background of the original visualization. Hint: Look for an image of a "parchment paper" on a free image site like Pixabay (https://pixabay.com) and place the image in the background of the plot.
9. Update the font to match, or closely resemble, the font used in the original visualization. Hint: Use the **showtext** package.

Activity 2

The second activity focuses on recreating Figure 1b, which 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.

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.

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

Then, we dive into the reproduction exercises, using the following dataset.

class	average_income	Rent	Food	Clothes	Tax	Other
$100-200	139.10	19	43	28	9.9	0.1
$200-300	249.45	22	47	23	4.0	4.0
$300-400	335.66	23	43	18	4.5	11.5
$400-500	433.82	18	37	15	5.5	24.5
$500-750	547.00	13	31	17	5.0	34.0
$750-1000	880.00	0	37	19	8.0	36.0
$1000 AND OVER	1125.00	0	29	16	4.5	50.5

1. Create the segmented bar plot. Hint: You will first need to pivot your data longer.
2. Add annotations that show the percentages associated with each of the segments.
3. Update the color palette and legend to match the original Du Bois visualization.
4. Re-position the legend and update axis labels and plot title.

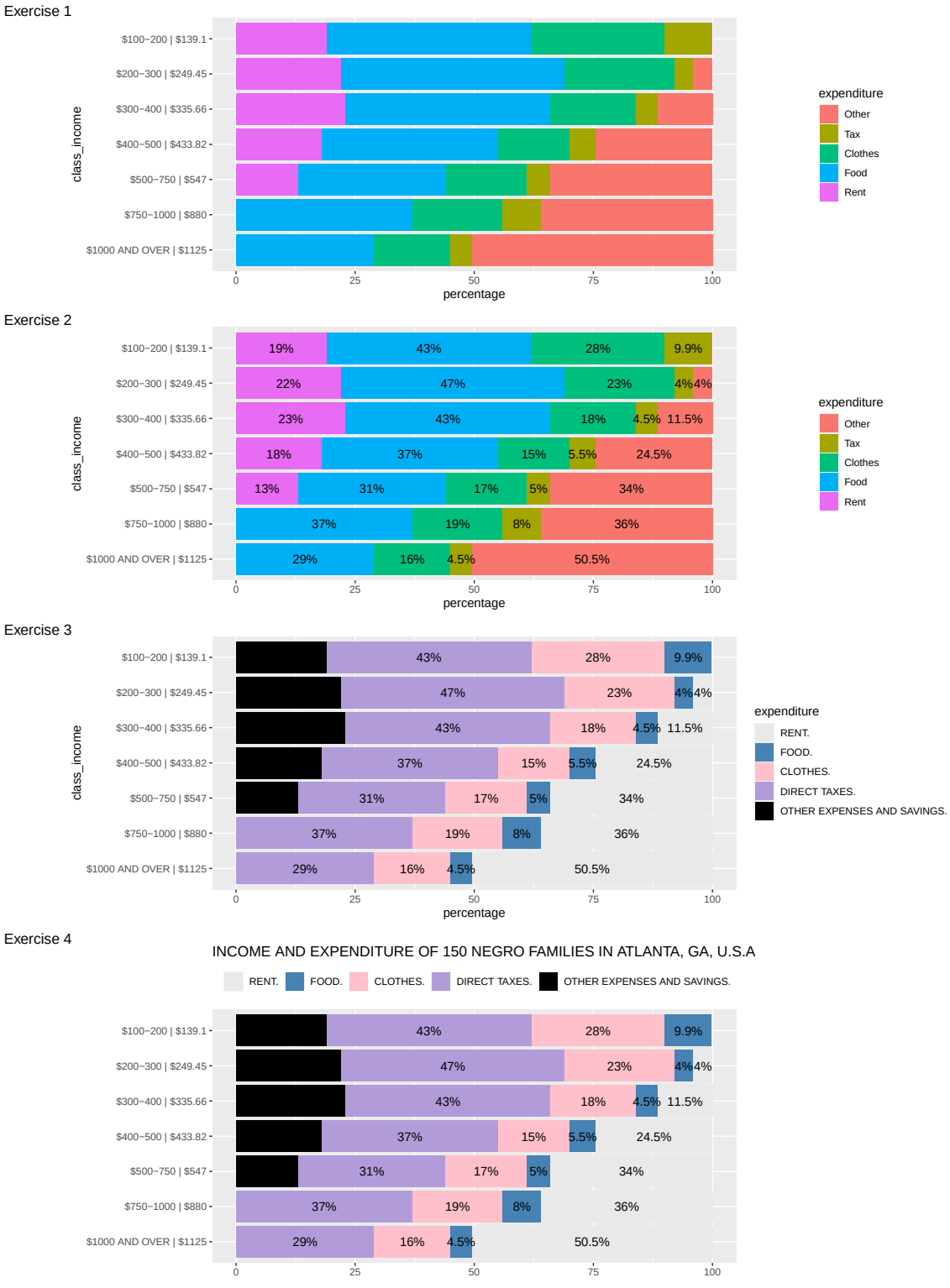
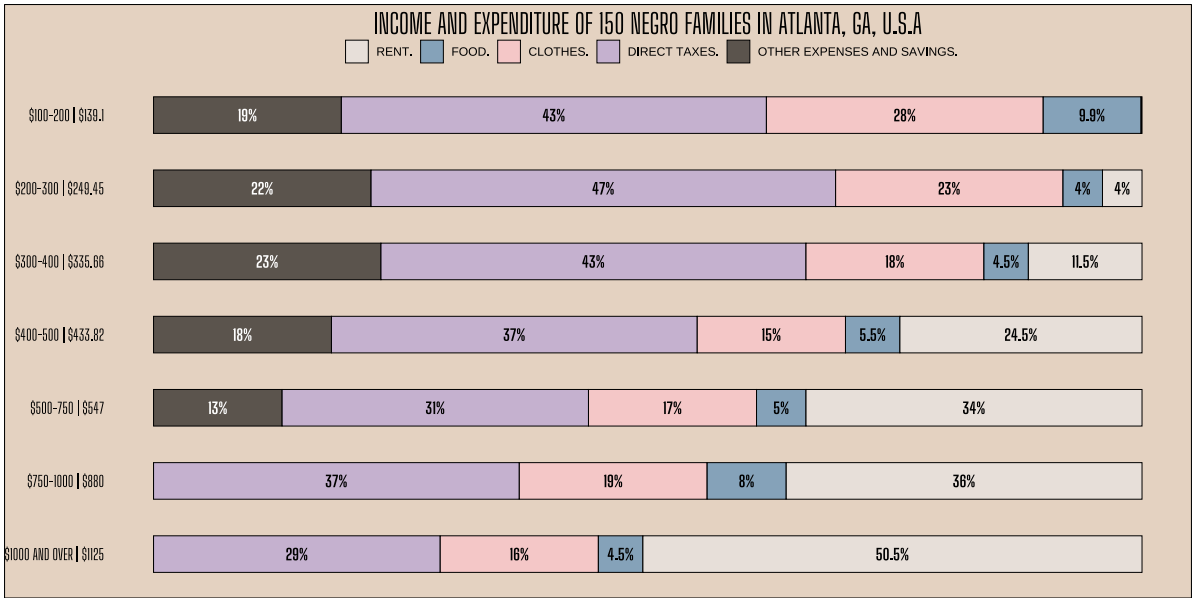


Figure 3: Figures to recreate for the exercises in Activity 2 - Part 1.

We still have quite a bit more modifications to make to get closer to recreating the original visualization. Part 2 of this activity requires pushing students' data visualization skills a bit further, even if it might not be feasible to achieve full reproduction of the original.

5. Recreate the following visualization. Hint: Use the Big Shoulders Display font, which you can download at https://fonts.google.com/specimen/Big+Shoulders+Display.



Conclusion

In this column we have presented two classrooms activities inspired by the #DuBoisChallenge as well as provided some background for W.E.B. Du Bois's work. These activities can be used in the classroom as is or serve as inspiration for recreating other Du Bois visualizations.

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
- 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
- Starks, A. (2022, February 1). The #DuBois Challenge. Nightingale. nightingaledvs.com/the-dubois-challenge