Working title: DuBois Challenge

Introduction

- Including data visualization history in statistics classes
- W.E.B. Du Bois contribution to data visualization
- Introduce Du Bois challenge
- Share two examples activities to use the challenges in an introductory and more advanced statistics class

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 chart that shows percentage of school aged Black children who are and are not enrolled public schools in the years 1876, 1886, and 1896.

For the first activity, which is aimed at the introductory statistics and data science audiences that teaches data visualization as part of its curriculum, will focus on getting a stacked bar chart representing the correct percentages. The second activity, aimed at a more advanced audience in a data visualization course, will build on it to capture all features of the visualization, including varying sizes of bars as well as colors and fonts.

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.

Activity 1

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

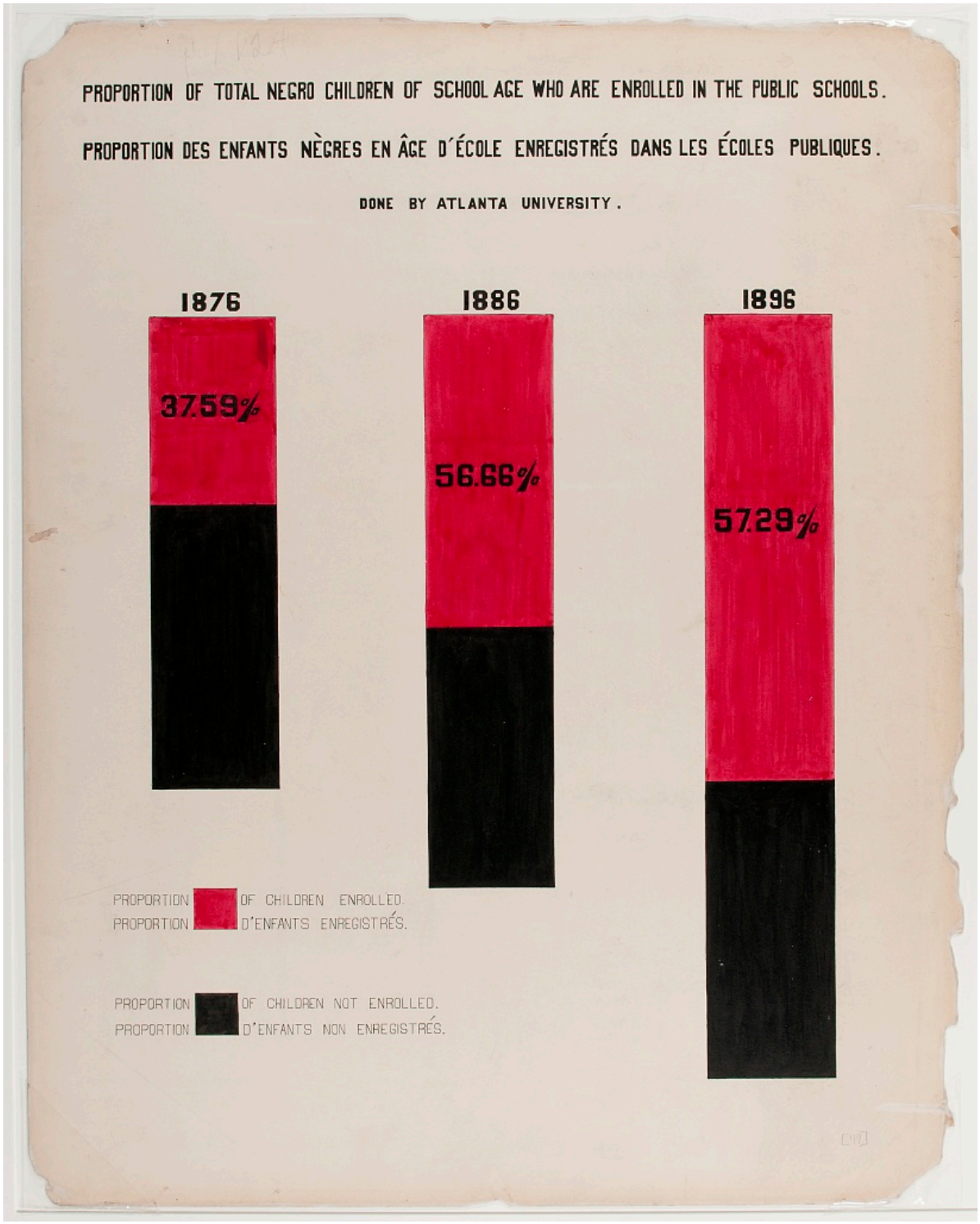


Figure 1: Proportion Of Total Negro Children Of School Age Who Are Enrolled In The Public Schools (plate 49)

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 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.

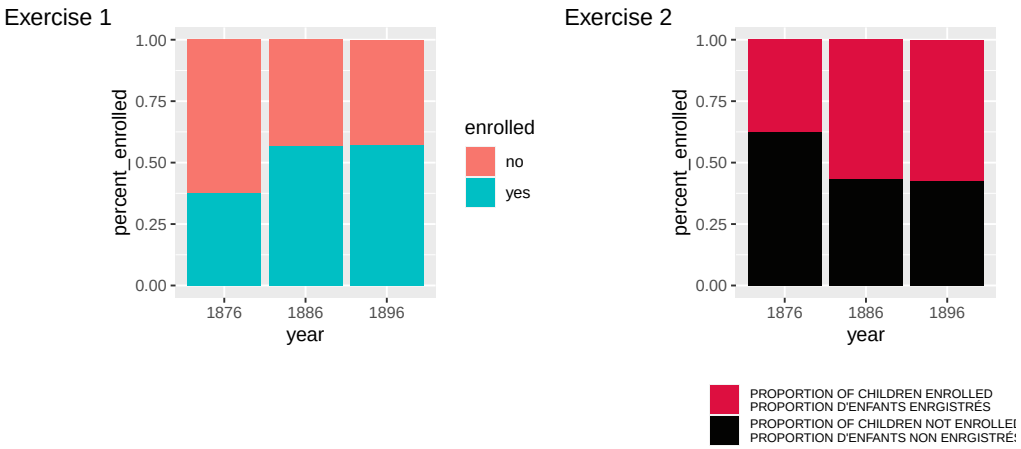
5. Match the plot title and subtitle

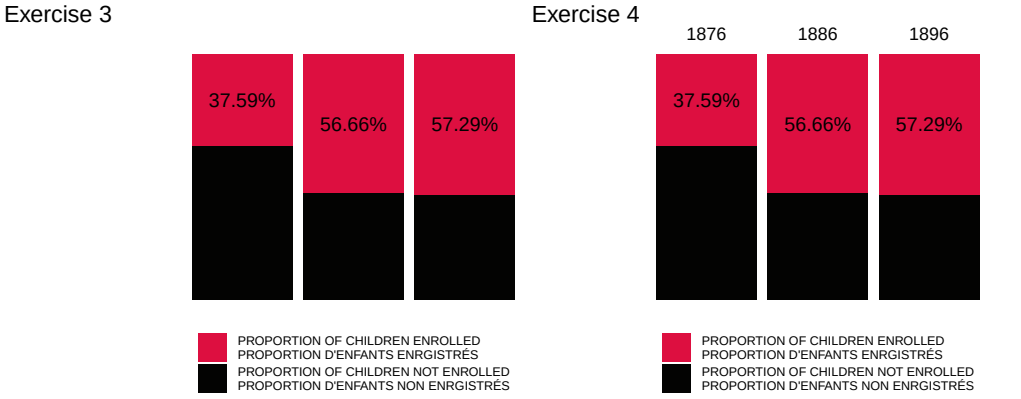
6. Make the bars skinnier and increase the white space between them.

Activity 2

Conclusion

Further reading





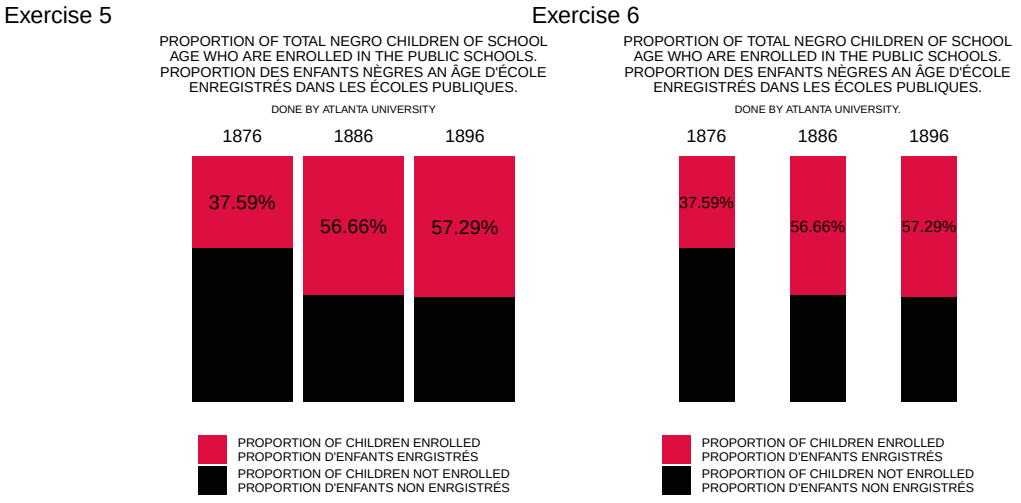


Figure 2: Expected result of each of the exercises in Activity 1.