Revisiting Du Bois

A Data Visualization Project

Nathan Kim

September 9th, 2021

Introduction

Prolific and prominent Black writer, historian, professor, political activist, W.E.B. Du Bois has left a legend in many ways.

This project has two conceptual goals.

1. **How does Du Bois' approach to data visualization depart from canonical views of data and information?**

In addition to being a sociologist who meticulously drew and recorded social statistics for over fifty years, Du Bois was also a Black theorist of knowledge, an educator, a novel writer, and a historian. Some of these perspectives are visible in his approach to data visualization. While other data visualization pioneers like Edward Tufte try to objectively gauge the value of a dataset as the proportion to which it represents "truth," Du Bois recognizes visualization as a creative endeavor alongside (quite literally, in the case of the Paris Exposition) photography and historical analysis. In extending Du Bois' data visualization through practical tools, I hope to also convey his artistic perspective on data visualization.

2. **How have the subjects of Du Bois' works in the 1900 Paris Exposition evolved over time?**

Du Bois was concerned with the "afterlives" of slavery and prospects of Black people in the American South after emancipation in 1865, and he provided answers by studing land ownership, occupations, income, and geographic concentration. How does the same topic of the afterlife of slavery in land ownership, occupations, income, and geographic dispersion look like in the contemporary age?

Project details

This project will consist of two components.

The first component is a ggplot2 package in R , tentatively called "ggdubois," to provide tools for Du Bois' data visualizations. The ggplot2 package implements a "grammar of graphics in R," creating an ecosystem of data visualization resources that expand off of a few simple functions. My package will provide an extension to this ecosystem by providing "geoms," or geometric representations of data, that emulate Du Bois' work. Some work has already been completed by others in the R ecosystem in this respect: the #tidytuesday challenge on Twitter has generated discussion and code for recreating Du Bois' symmetrical pie chart and an iconic spiral graph. Others have created a ggplot2 theme with a Du Bois-style color palette, viewable here. My contribution to this project would be to extend these to include Du Bois' wrapped bar chart (42), scalable map, woven bar chart (41), spike-embedded circle chart (44), meandering spiral chart (40),

"mountain peak" chart (52), and bidirectional bar chart (163). Also, while most discussion around Du Bois' "data portraits" have revolved around his work for the 1900 Paris Exposition, my work will also incorporate his graphics from other seminal works including *The Philadelphia Negro*, *Black Reconstruction*, *Color and Democracy*, and *The World and Africa* in the package.

The second component of my project will be a website extending Du Bois' explorations into the contemporary era. The website will display a series of comparisons of two visualizations side by side, one originally drawn by Du Bois and one extended to topics of modern day. As opposed to literally emulating Du Bois' topic and adjusting only the time frame, these reanlyses will also include contemporary tools of interactivity and animation. Statistical analyses will also be incorporated into this visualization; for example, hierarchical clustering and Moran's I tests will be used to analyze geogrpahic dispersion.

Data for this reanalysis will come from the American Community Survey (ACS) for contemporary figures and the Census Bureau's Public Use Microdata Sample (PUMS) files for time-series data. The visualizations on this website will be built through the ggdubois package and the d3.js library in JS.

The text for the end-of-term report for S&DS491 will largely come from the text in the website component.

GitHub repositories for the ggplot2 package and the website can be found here and here.

Timeline

October 1st: Prototype of ggplot2 functions ready. Datasets for reanalysis obtained and cleaned; statistical plan for reanalysis cleared.

October 15th: ggplot2 package is publication-ready (documentation with examples, vignettes, and auxilliary datasets included).

November 1st: Draft reanalysis of Du Bois' topics in the website component ready.

November 15th: Reanalysis revisions completed. Website composition and formatting begun.

December 1st: Draft of website is finished.

December 10th: Senior requirements (Poster, project report) finished; revisions to website finished