

Portfolio 1

Michael Kornely

2/20/2022

Why did you choose the data that you did? What makes it interesting to you?

I choose to use a public Madison Dataset (located here:<https://data-cityofmadison.opendata.arcgis.com>) that contains several different demographics for each one of its neighborhoods. I am not familiar with a lot of these neighborhoods, given that I spend a majority of my time mostly on the Campus and State Street areas. I wanted to understand the racial makeup, as well as connect that to possible economic inequality, food deserts, and kids struggling in school. Using these demographics, we may be able to uncover systemic issues in Madison Neighborhoods and find out ways to try to alleviate those problems. I hope to look at other Madison datasets to understand the city I live in more clearly and help other students understand the community they are a part of.

Figure A: Madison Neighborhoods by Racial Makeup

```
madison%>%select(name,pc_wht_10,pc_asn_10,pc_afrm_10,pc_othm_10,pc_hisp_10)%>%
  filter(name!="City of Madison")%>%
  rename(White=pc_wht_10,Asian=pc_asn_10,
         "African-American"=pc_afrm_10,
         Other=pc_othm_10,
         Hispanic=pc_hisp_10)%>%
  pivot_longer(cols=White:Hispanic,names_to="Race",
               values_to = "Proportion") %>%
  ggplot(aes(x=Race, y=Proportion,fill=Race))+
  geom_col() + facet_wrap(~name)+
  coord_flip()+ylab("Proportion of Total Citizens")
```

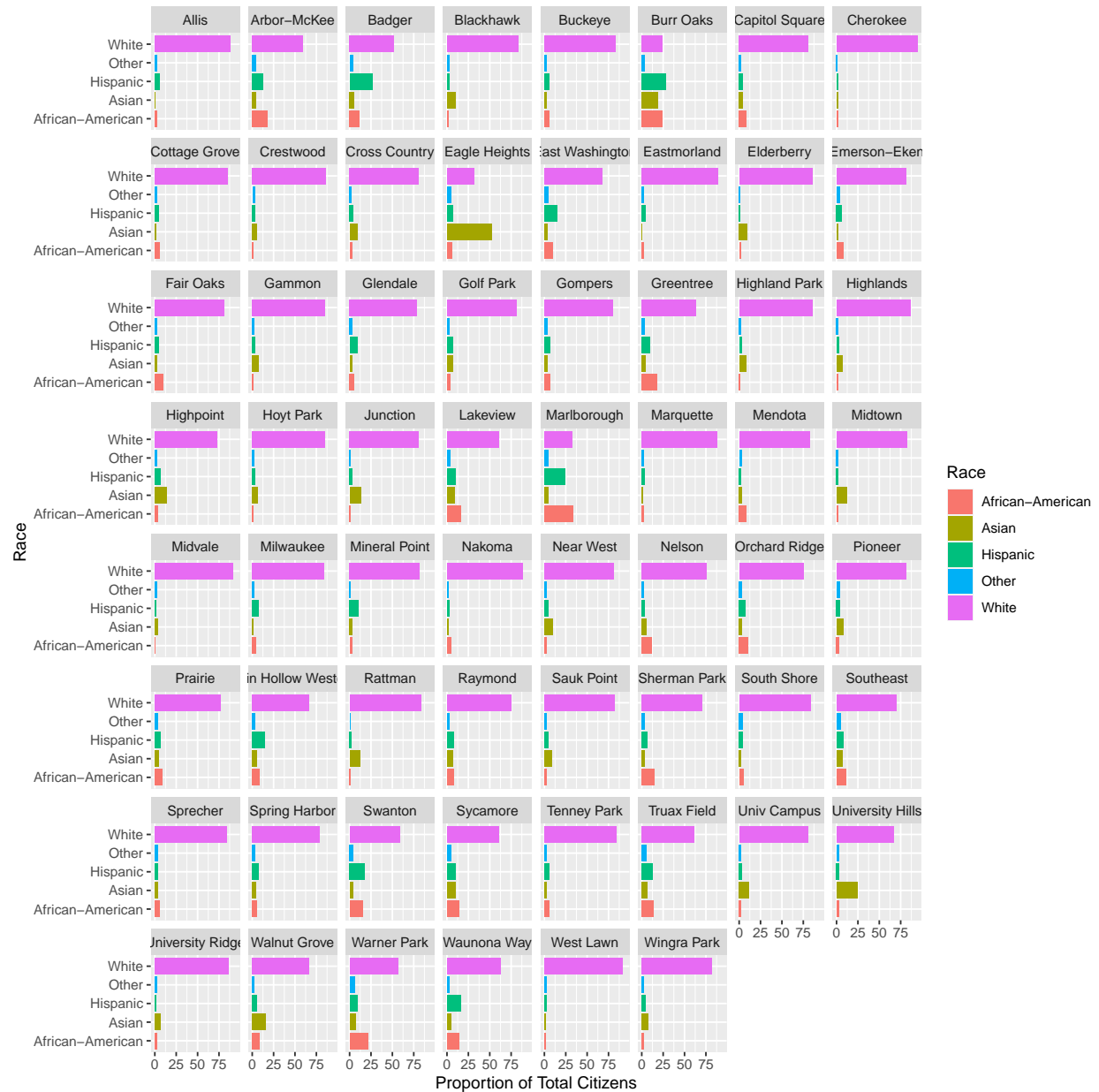
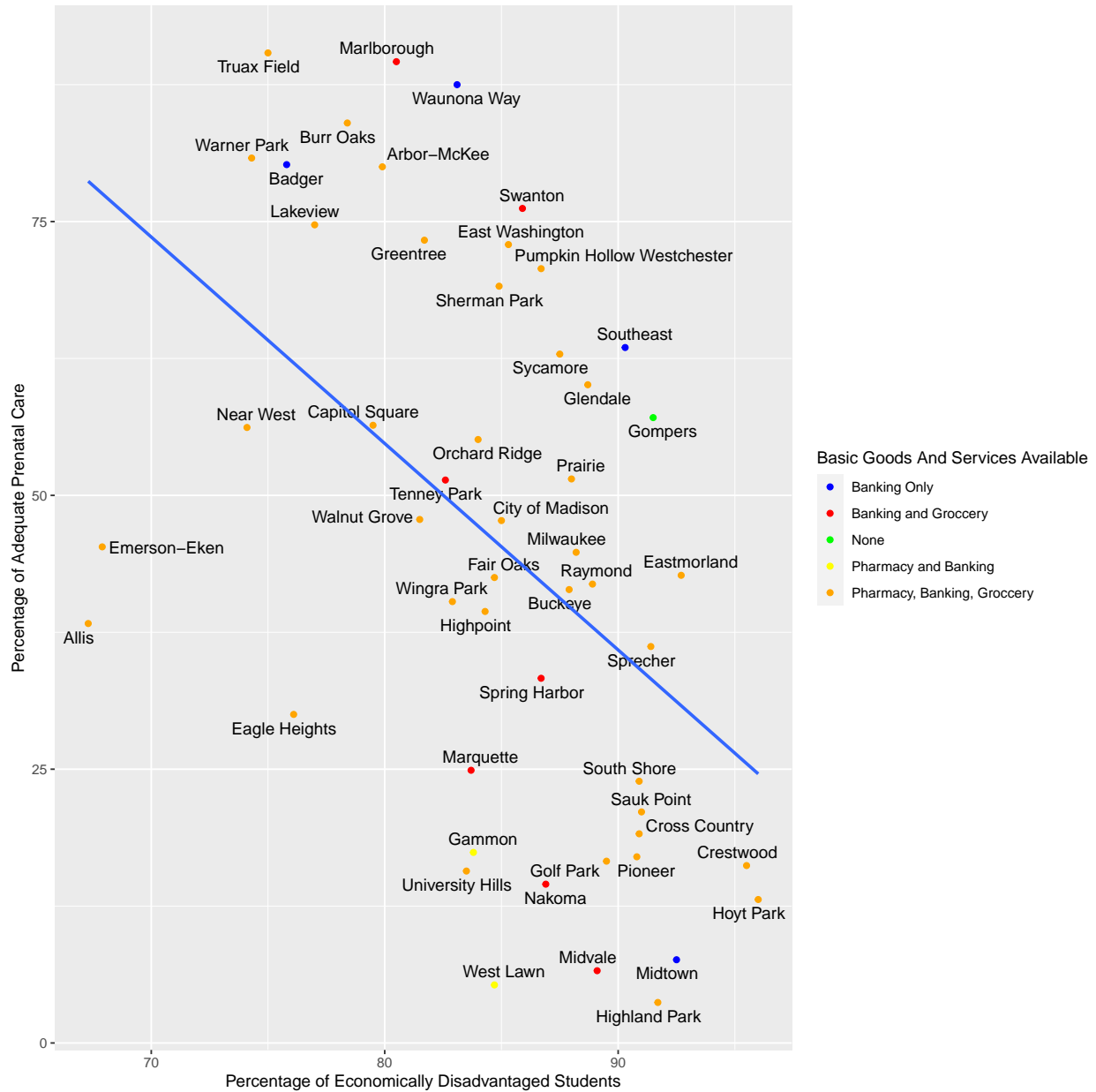


Figure B: Percentage of Economically Disadvantaged Students vs Percentage of Adequate Prenatal Care

```
madison%>%select(name, pc_frlu, pc_appr,gds_svcs)%>%
  filter(pc_frlu!="suppressed")%>%
  filter(pc_appr!="suppressed")%>%
  mutate(pc_frlu=as.numeric(pc_frlu))%>%
  mutate(pc_appr=as.numeric(pc_appr))%>%
  ggplot(aes(y=pc_frlu,x=pc_appr))+geom_point(aes(color=gds_svcs))+
  geom_text_repel(aes(label = name))+
  geom_smooth(method='lm',se=FALSE) +
```

```
xlab("Percentage of Economically Disadvantaged Students")+
ylab("Percentage of Adequate Prenatal Care")+
labs(color='Basic Goods And Services Available') +
scale_color_manual(labels = c("Banking Only", "Banking and Grocery",
                             "None",
                             "Pharmacy and Banking", "Pharmacy, Banking, Grocery"),
                  values = c("blue", "red", "green", "yellow", "orange"))
```

```
## 'geom_smooth()' using formula 'y ~ x'
```



What are some interesting facts that you learned through the visualizations? Provide at least one unexpected finding.

Some interesting facts that I found were that a majority of the neighborhoods seemed to be predominantly white. Given that the city in itself is a majority white, that was not surprising, however, I believed there may be more minority-majority areas in the city like what we see in other cities (i.e Chinatown). Most neighborhoods seemed to have the same general makeup as the rest of the city, with a few outliers where White individuals made up a plurality

That led me to my next visualization where I choose to make a relationship between Prenatal Care and Economically Disadvantaged Students. I was not very surprised to find evidence of a linear relationship, however, it seems that the Basic Goods and Services that were available show any clear patterns. There were several more relationships that I would like to look into, however, none were qualitative to include in this visualization.

How did you create the plot? Were there any data preparation steps you used; e.g.,

using `pivot_longer` or `mutate`? What guided the style customizations you used?

When I created my initial plot, I pivoted longer to be able to facet the data by each neighborhood and create the visual that I wanted to as each race variable was its column. I also flipped the graph to make it more readable and edited some column names so they make sense to the reader.

In the second graph, I had to clean up the data a bit as the observations were characters instead of integers. I also edited several of the labels to make it more readable and used `geom_text_repel` to label each point and be able to connect it to my earlier visualization to see if there were patterns with race as well. I used a linear regression line to show the possible relationship between the two variables to make sense of the scatter plot.