PHW251 Data Project Milestone 4

Nakisa Golabi, Joanna Liang, Desmond Gill

11/16/2022

Table 1: Average Pack-Years Per Disease Outcome

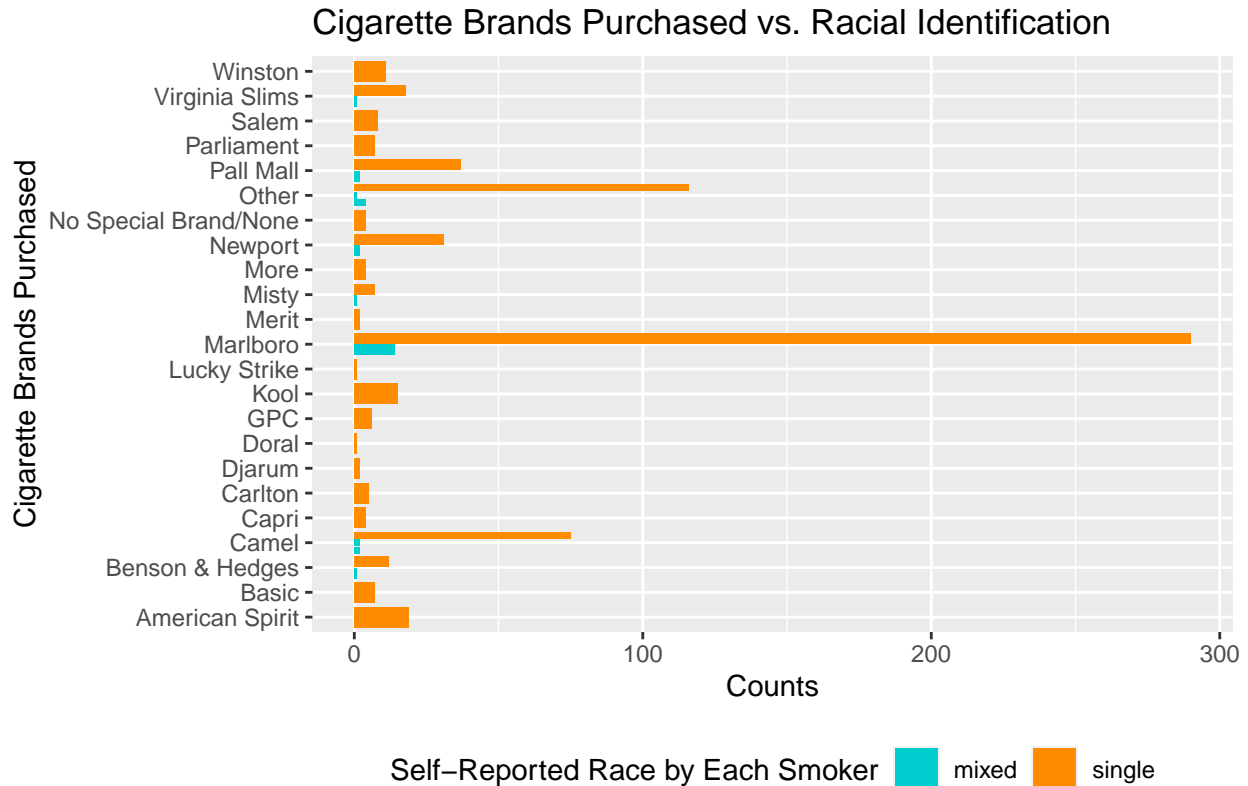| Outcomes	Pack-Years
Asthma	23.93
Heart Disease	28.83
Mental Illness	18.94
Diabetes	30.10

PRINT-QUALITY TABLE THAT SHOWS THE AVERAGE NUMBER OF PACK-YEARS AND THE DISEASE OUTCOMES

```
kable(tibble_py_disease, booktabs = T, digits = 2, escape=F,
                 col.names=c("Outcomes", "Pack-Years"),
                 align="cc",
                 caption = "Average Pack-Years Per Disease Outcome")
```

The table compares the average number of pack-years (the product of the number of packs of cigarettes smoked per day and the years a person has smoked) by the disease outcomes of asthma, heart disease, diabetes, and mental illness. Those with diabetes had the highest mean and those with mental illness had the lowest mean.

```
ggplot(data=df_rb, aes(x = counts, y=factor(smkbrand), fill=race_binary)) +
  geom_bar(aes(fill=race_binary), stat="identity", position=position_dodge2()) +
  labs(x="Counts",y="Cigarette Brands Purchased",
       title="Cigarette Brands Purchased vs. Racial Identification",
       caption="Data Source: CDPH 2011 California Smokers Cohort.")+
  scale_fill_manual(name="Self-Reported Race by Each Smoker",
                    values=c("cyan3","darkorange", "darkgreen"))+
  theme(legend.position="bottom")
```

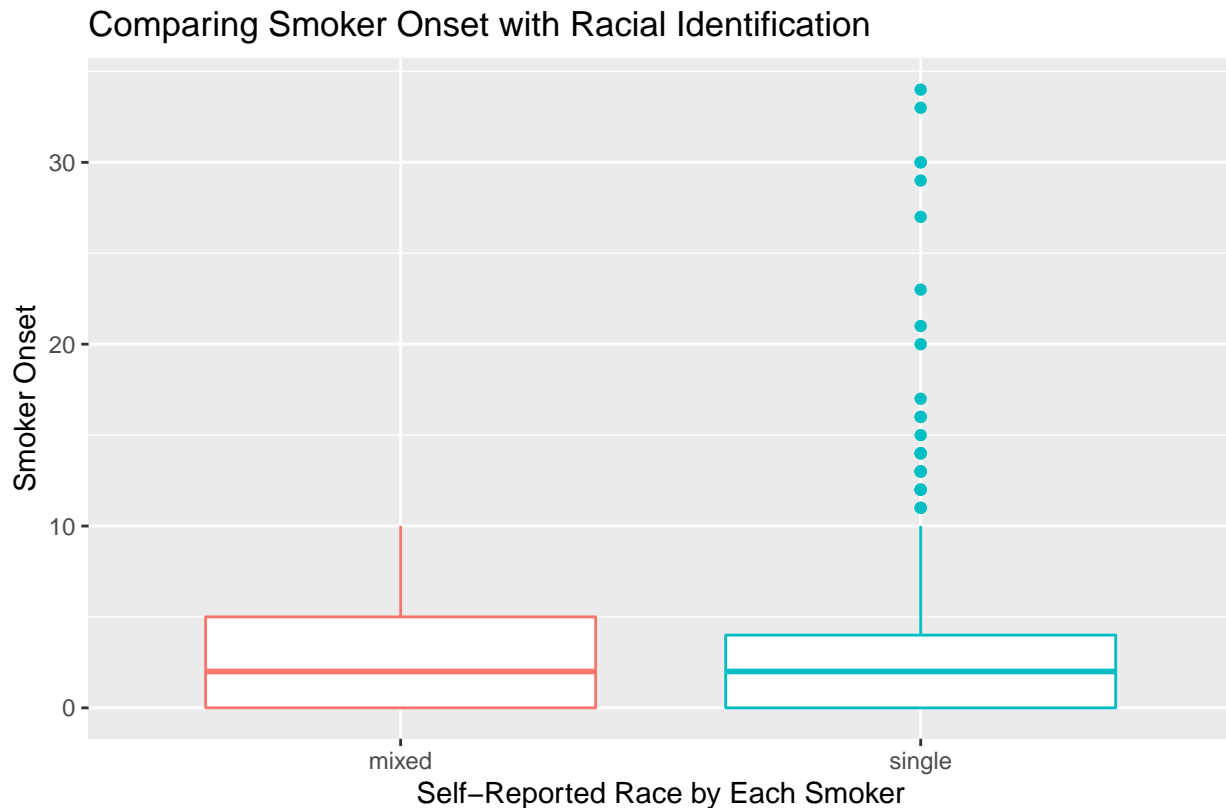## Cigarette Brands Purchased vs. Racial Identification

Data Source: CDPH 2011 California Smokers Cohort.

This visualization compares the cigarette brands purchased by each smoker with their racial identification (self-reported mixed or single race), in the CDPH 2011 California Smokers Cohort. By knowing that out of 712 total smokers, Marlboro was the highest purchased cigarette brand for both self-reported mixed race smokers (at 14 counts) and self-reported single race smokers (at 290 counts), we can target this specific cigarette brand to help reduce certain disease outcomes (heart disease, asthma, mental illness, diabetes).

```
ggplot(data=df_rs, aes(x=race_binary,y=smoker_onset)) +
  geom_boxplot(aes(colour=race_binary)) +
  theme(legend.position="none") +
  labs(x="Self-Reported Race by Each Smoker",y="Smoker Onset",
       title="Comparing Smoker Onset with Racial Identification",
       caption= "Data Source: CDPH 2011 California Smokers Cohort.")
```


Comparing Smoker Onset with Racial Identification

Data Source: CDPH 2011 California Smokers Cohort.

This box plot compares smoker onset (the difference between the age when they started smoking regularly and the age of the participant's first cigarette) with their racial identification (self-reported mixed or single race), in the CDPH 2011 California Smokers Cohort. The medians for smoker onset were 2.00 for smokers with self-reported mixed and 2.00 for single race, thus further analysis into additional behavioral factors beyond racial identification is needed to help reduce certain disease outcomes (heart disease, asthma, mental illness, diabetes).