

Cap DAP Final Copy

Aubrey Ahern

12/14/2022

Introduction

- I took my data from <https://data.chhs.ca.gov/dataset/hiv-aids-cases>.
- The data was collected from 2011-2017 within the state of California in order to collect public health information and create policy on HIV/AIDS within California.
- The data outlines all cause deaths in HIV/AIDS by gender, age, race/ethnicity, and transmission category.
- I used this data to take a look at the distribution of HIV/AIDS cause deaths by different demographic variables, primarily looking at how good of a predictor gender is in determining health outcomes.
- While there is no biologic reasoning for why certain groups of people experience HIV/AIDS cause deaths at a higher rate, we can call upon public health knowledge to answer these questions.
 - When taking social determinants of health into consideration, we may see that women and LGBTQ+ communities experience lower accessibility to healthcare and are often subjected to discrimination within the healthcare field. This provides us with a brief explanation of the data that will be presented throughout this project (Sharman).

Getting Started

- I made sure my data was compatible to work with R by...
 - Excluding spaces and special characters from column names
 - Organizing categories into columns instead of rows
 - Making sure that each cell contained only one piece of information.
- After importing my data, I made all variables factors, except for count which I made numeric.

Limitations

- While this data set contains a lot of information on demographic information on those who have died from HIV/AIDS, it only includes **one** piece of information per row, instead of multiple per datum. I have worked with the data as best I could but would potentially use a more extensive data set in the future.

Analysis

Test #1 - Analyzing the relationship between gender and cause deaths by HIV/AIDS.

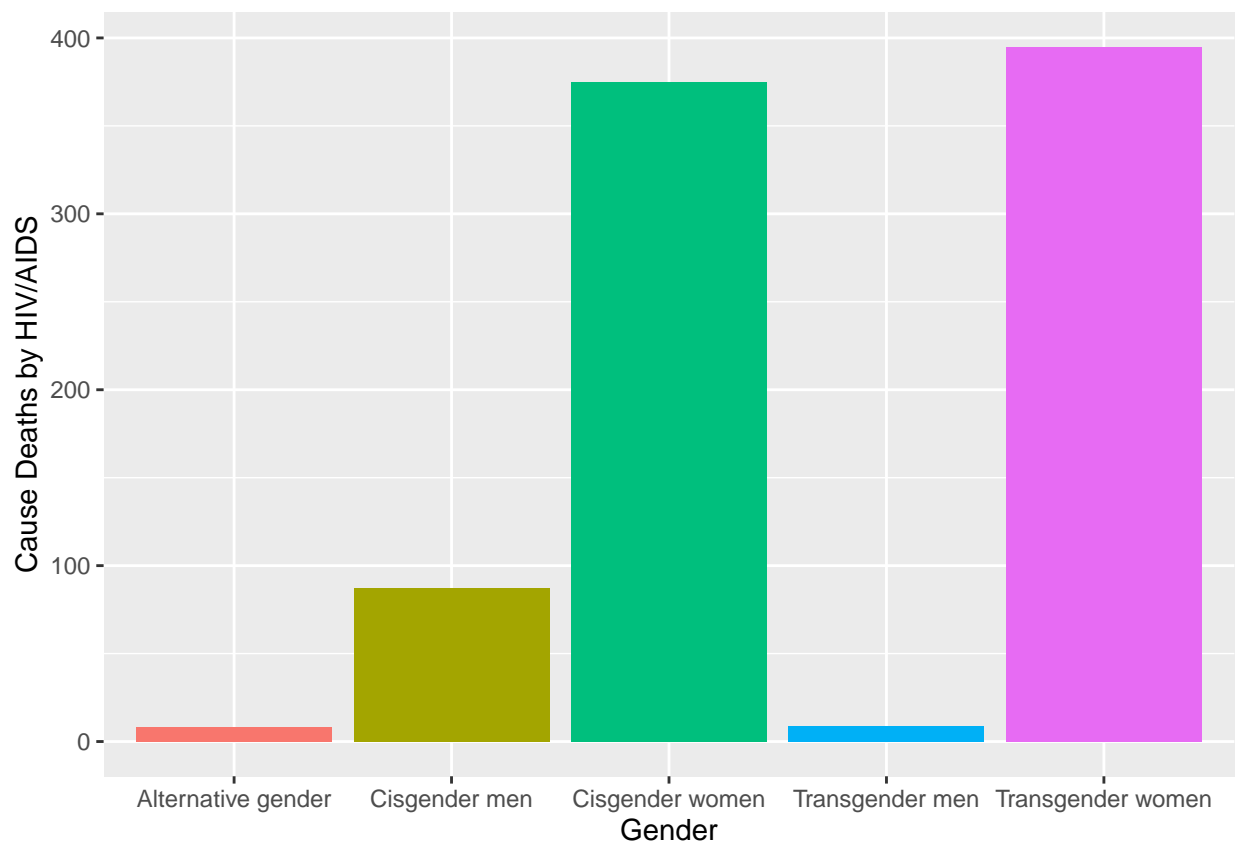
- First, I created a data frame containing only the data I wanted.

```
DF_gender <- DF %>%
  group_by(current_gender) %>%
  summarise(count = sum(count)) %>%
  arrange(desc(count)) %>%
  na.omit()
```

DF_gender

- Next, let's visualize this relationship.

```
ggplot(DF_gender, aes(current_gender, count, fill = current_gender))+
  geom_col()+
  labs(x = "Gender",
       y = "Cause Deaths by HIV/AIDS") +
  theme(legend.position = "none")
```



Hypothesis: Trans women and cis women experience HIV/AIDS cause deaths at a higher rate than those of other genders.

Null Hypothesis: There is no correlation between gender and HIV/AIDS cause deaths.

- I then ran a statistical test to see if I can accept or reject it. This relationship includes one piece of categorical data (current gender) and one piece of count data (count). While I originally decided to run an ANOVA test, a tukey test would be more appropriate as it will assess the variability between every comparison of two genders... so let's do it.

- These results indicate the following conclusions:
 - There is no significant difference in the number of cause deaths by HIV/AIDS between cisgender men and alternative gender people.
 - **Cisgender women experience significantly more cause deaths by HIV/AIDS than alternative gender people.**
 - There is no significant difference in the number of cause deaths by HIV/AIDS between transgender men and alternative gender people.
 - **Transgender women experience significantly more cause deaths by HIV/AIDS than alternative gender people.**
 - **Cisgender women experience significantly more cause deaths by HIV/AIDS than cisgender men.**
 - There is no significant difference in the number of cause deaths by HIV/AIDS between transgender men and cisgender men.
 - **Transgender women experience significantly more cause deaths by HIV/AIDS than cisgender men.**
 - **Cisgender women experience significantly more cause deaths by HIV/AIDS than transgender men.**
 - There is no significant difference in the number of cause deaths by HIV/AIDS between transgender women and cisgender women.
 - **Transgender women experience significantly more cause deaths by HIV/AIDS than transgender men.**
- Thus, both cisgender and transgender women experience significantly more cause deaths by HIV/AIDS than those of any other gender but do not experience significantly different rates of cause deaths by HIV/AIDS from one another and our null hypothesis may be rejected.

Test #2

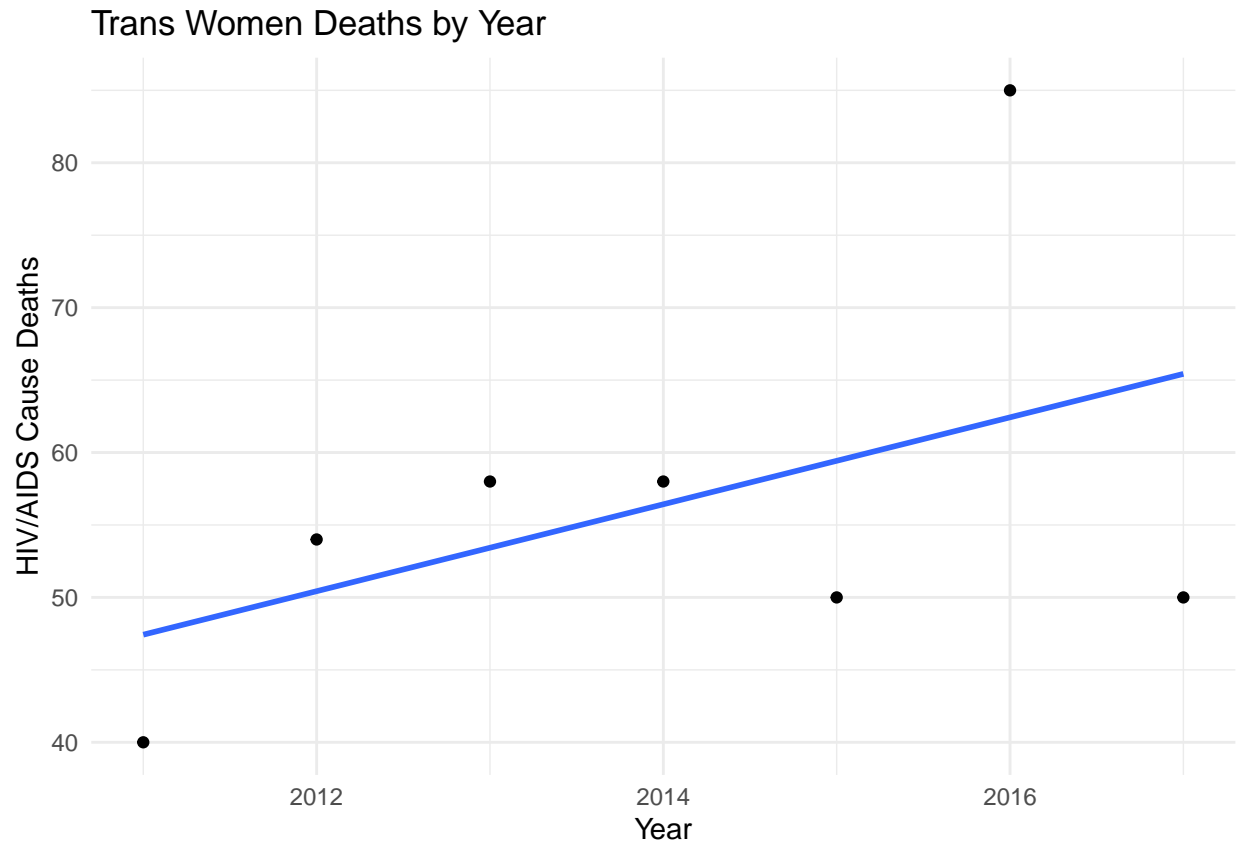
- Since we now know that trans and cis women experience more deaths by HIV-AIDS than those of other genders, let's analyze the relationship between year and deaths for trans and cis women.
- I created a new data frame that only contains data for trans women.

```
DF_transwomen <- DF %>%
  filter(current_gender == "Transgender women")%>%
  mutate(year = as.numeric(year),
         count = as.numeric(count))
DF_transwomen
```

- Then I visualized this new dataframe.

```
ggplot(DF_transwomen, aes(x = year, y = count)) +
  geom_point() +
  geom_smooth(method = "lm", se = FALSE)+
  labs(x = "Year",
       y = "HIV/AIDS Cause Deaths",
       title = "Trans Women Deaths by Year") +
  theme_minimal()
```

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



Hypothesis: Cause deaths by HIV/AIDS increase over time for transwomen

Null Hypothesis: There is no correlation between HIV/AIDS cause deaths for trans women by year.

- Then, I ran a t-test to see if there is a relationship between time (in years) and death count for trans women.
 - I chose to conduct a t-test rather than another ANOVA test as I considered year to be numeric data rather than categorical.

```
myMod2 <- lm(count ~ year, data = DF_transwomen)
summary(myMod2)
```

```
##
## Call:
## lm(formula = count ~ year, data = DF_transwomen)
##
## Residuals:
##      1      2      3      4      5      6      7
## -7.429  3.571  4.571  1.571 -9.429 22.571 -15.429
```

```
##
## Coefficients:
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept) -5985.571   5184.451  -1.155   0.300
## year          3.000     2.574   1.165   0.296
##
## Residual standard error: 13.62 on 5 degrees of freedom
## Multiple R-squared:  0.2136, Adjusted R-squared:  0.05633
## F-statistic: 1.358 on 1 and 5 DF,  p-value: 0.2964
```

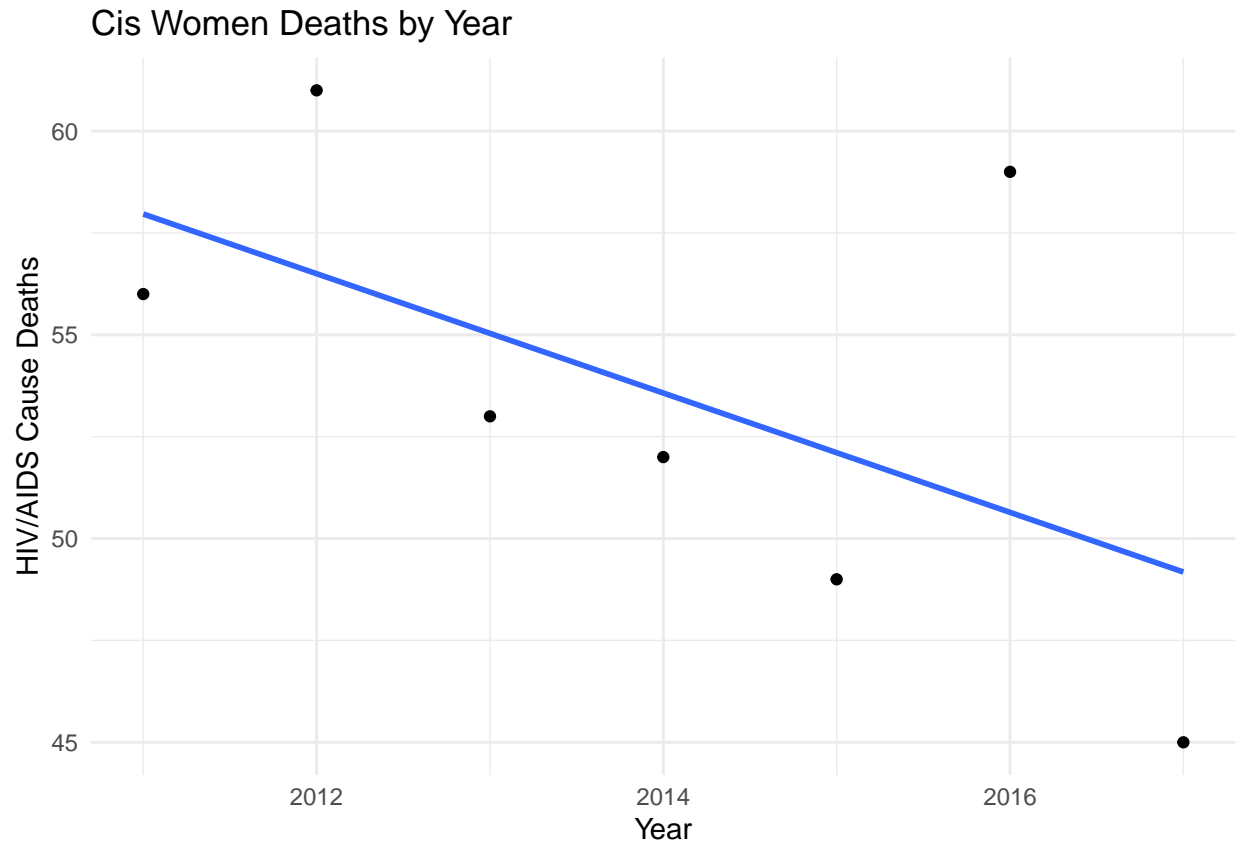
- This allowed me to accept the null hypothesis that there was no observable relationship between year and death count for trans women as the p value, $0.2964 > 0.05$. Below is a figure which represents the (lack of) correlation within this data.
- I then did the same for cis women by first creating a new data frame that only contains the data I wish to analyze.

```
DF_ciswomen <- DF %>%
  filter(current_gender == "Cisgender women")%>%
  mutate(year = as.numeric(year),
         count = as.numeric(count))
DF_ciswomen
```

- Again, I visualized this new data frame...

```
ggplot(DF_ciswomen, aes(x = year, y = count)) +
  geom_point() +
  geom_smooth(method = "lm", se = FALSE)+
  labs(x = "Year",
       y= "HIV/AIDS Cause Deaths",
       title = "Cis Women Deaths by Year") +
  theme_minimal()
```

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



Hypothesis: Cause deaths by HIV/AIDS decrease over time for cis women.

Null Hypothesis: There is no correlation between HIV/AIDS cause deaths for cis women by year.

```
myMod3 <- lm(count ~ year, data = DF_ciswomen)
summary(myMod3)
```

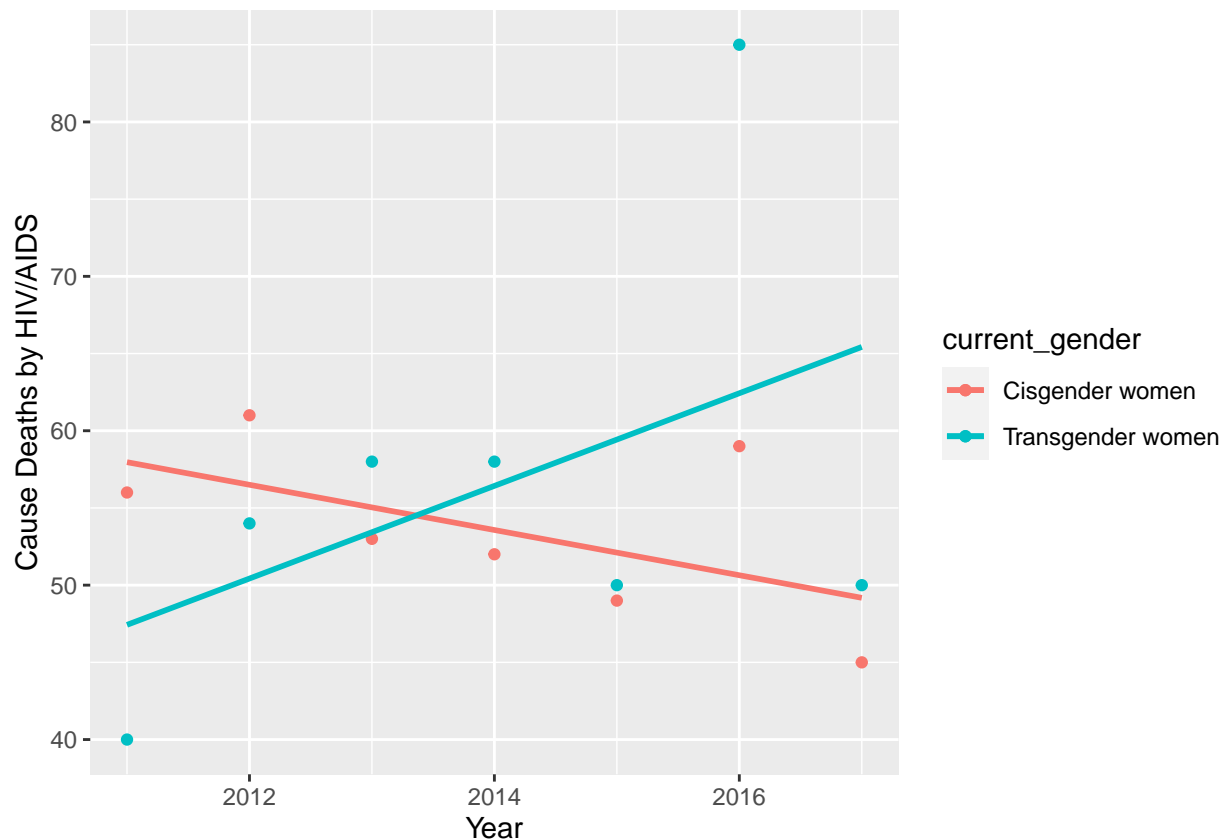
```
##
## Call:
## lm(formula = count ~ year, data = DF_ciswomen)
##
## Residuals:
##      1      2      3      4      5      6      7
## -1.964  4.500 -2.036 -1.571 -3.107  8.357 -4.179
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3002.643   1923.334   1.561   0.179
```

```
## year          -1.464          0.955  -1.533    0.186
##
## Residual standard error: 5.053 on 5 degrees of freedom
## Multiple R-squared:  0.3198, Adjusted R-squared:  0.1838
## F-statistic: 2.351 on 1 and 5 DF,  p-value: 0.1858
```

- This allowed me to accept the null hypothesis that there was no observable relationship between year and death count for cis women as the p value, $0.1858 > 0.05$. Below is a figure which represents the (lack of) correlation within this data. Here is both data plotted on the same graph in summary.

```
ggplot(data = DF_ciswomen, aes(x = year, y = count, colour = current_gender)) +
  geom_point() +
  geom_smooth(method = "lm", se = FALSE) +
  geom_point(data = DF_transwomen, aes(x = year, y = count, color = current_gender)) +
  geom_smooth(data = DF_transwomen, aes(x = year, y = count, color = current_gender), method = "lm", se = FALSE) +
  labs(x = "Year",
       y = "Cause Deaths by HIV/AIDS")
```

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



Test #3

- Lastly, I wanted to consider which category, race, gender, age, or transmission method, is the best predictor of cause deaths by HIV/AIDS. The data I have selected however, does not lend itself to this nicely. Thus, I chose to analyze the relationship between race and cause deaths by HIV/AIDS.
- First, I created my new data frame.

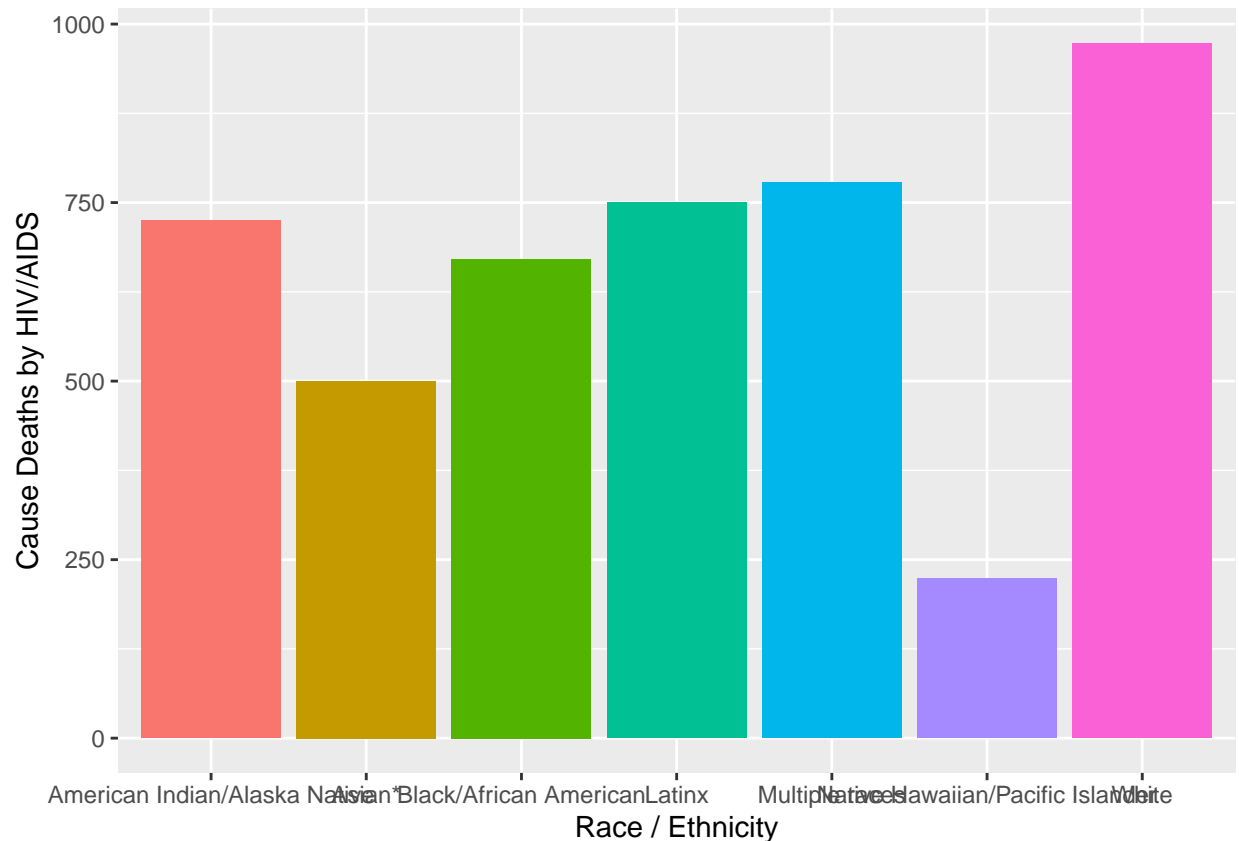
```
DF_race_ethnicity <- DF %>%  
  group_by(race_ethnicity) %>%  
  summarise(count = sum(count)) %>%  
  arrange(desc(count)) %>%  
  na.omit()
```

```
DF_race_ethnicity
```

```
## # A tibble: 7 x 2  
##   race_ethnicity      count  
##   <fct>           <dbl>  
## 1 White              973  
## 2 Multiple races     778  
## 3 Latinx            750  
## 4 American Indian/Alaska Native 725  
## 5 Black/African American 671  
## 6 Asian*            500  
## 7 Native Hawaiian/Pacific Islander 224
```

- Then, I visualized this relationship...

```
ggplot(DF_race_ethnicity, aes(race_ethnicity, count, fill = race_ethnicity))+  
  geom_col()+  
  labs(x = "Race / Ethnicity",  
       y = "Cause Deaths by HIV/AIDS")+  
  theme(legend.position = "none")
```

Hypothesis: Those who are white experience more cause deaths by HIV/AIDS than those of other races and ethnicities.

Null Hypothesis: There is no relationship between race/ethnicity and cause deaths by HIV/AIDS.

- Now, let's run another tukey test to see if this is true. . .
- These results indicate the following conclusions:
 - Native Hawaiian/Pacific Islander people experience significantly **less** cause deaths by HIV/AIDS than those who are American Indian/Alaskan Native.
 - Native Hawaiian/Pacific Islander people experience significantly **less** cause deaths by HIV/AIDS than those who are Black/African American.
 - Native Hawaiian/Pacific Islander people experience significantly **less** cause deaths by HIV/AIDS than those who are Latinx.
 - Native Hawaiian/Pacific Islander people experience significantly **less** cause deaths by HIV/AIDS than those who are multiple races.
 - White people experience significantly **more** cause deaths by HIV/AIDS than those who are Asian.

- White people experience significantly **more** cause deaths by HIV/AIDS than those who are Native Hawaiian/Pacific Islander.

Biological Summary

- There is no biological reasoning for why certain genders experience HIV/AIDS at higher rates than others aside from the fact that the virus is present in their communities.
- As seen throughout the analysis portion of this project, gender is the best categorical predictor of HIV/AIDS cause deaths but this does not mean in any way that gender causes HIV/AIDS. Instead, the current social order discriminates against certain groups such as trans folks and women based on their gender. The medical field is no exception from this social order and thus, these groups struggle to access equitable health care and therefore experience worse health outcomes (Sharman).
- People who are within marginalized communities often experience instances where physicians minimize their pain and ignore important health indicators such as the ones observed in HIV/AIDS. This can lead to misdiagnoses as well as late diagnoses (Kronenfeld). Catching HIV/AIDS within its early stages is crucial in dictating how the disease progresses. Thus, this may be a reason why trans women and cis women are more likely to experience HIV/AIDS cause deaths.
- This idea is challenged within my third test as it shows white people as being the most susceptible to cause deaths by HIV/AIDS. The data displayed in Test #3 however may be explained by geography. Here, we can see that Native Hawaiiin/Pacific Islander people experienced the lowest number of cause deaths by HIV/AIDS. This could be attributed to the fact that these islands are fairly isolated and would require a lot of travel in order to bring the HIV/AIDS disease to the islands.
- While many sources state that white people have better access to health care than any other racial group, this may skew the data when we think about data collection (Kronenfeld). Because white people generally have better access to healthcare, cases of HIV/AIDS may simply be **recognized, reported, and treated** at a higher rate for white people than for people of other races and ethnicities.

Challenges

- The main take away from this analysis is that gender is a key predictor for HIV/AIDS cause deaths and that trans and cis women are more likely to experience death due to HIV/AIDS.
- After completing this project, there is a few things I would do differently...
 - I would spend more time considering which data I would use. The data I found was pretty difficult to work with and didn't contain as much information as I had hoped.
 - I would find data with a larger sample size over a greater amount of time.
 - * Certain analyses could have been improved if my data spanned over more years.
 - I would conduct more tests if possible.

References

Zena Sharman. (2016). The Remedy : Queer and Trans Voices on Health and Health Care. Arsenal Pulp Press.

Jennie Jacobs Kronenfeld. (2018). Gender, Women's Health Care Concerns and Other Social Factors in Health and Health Care: Vol. First edition. Emerald Publishing Limited