

Final Report

due November 16, 2021 by 11:59 PM

Team TBD: Maggie Lundberg, Riya Mohan, and Izzy Kjaerulff

11/16/2021

Reading Data and Data Clean Up:

```
unzip("../data/IHME_DEX_ED_SPENDING_2006_2016_DATA.zip", exdir = "../data/spending_data_unzip")

spending <- read.csv("../data/spending_data_unzip/IHME_DEX_ED_SPENDING_2006_2016_DATA_Y2021M09D23.CSV")

library(tidyverse)
library(tidymodels)
```

Does the emergency department spend a different amount of money on males and females? This is looking at all spending, not taking into account type of insurance.

!! there was a both category i filtered out bc not sure what to do with ?

```
spending_malefemale <- spending %>%
  filter(sex %in% c("Female", "Male"))

t.test(spending_malefemale$mean_all~spending_malefemale$sex) %>%
  print()
```

```
##
## Welch Two Sample t-test
##
## data: spending_malefemale$mean_all by spending_malefemale$sex
## t = 4.0269, df = 6416.2, p-value = 5.717e-05
## alternative hypothesis: true difference in means between group Female and group Male is not equal to
## 95 percent confidence interval:
## 56638431 164092573
## sample estimates:
## mean in group Female mean in group Male
## 413610916 303245414
```

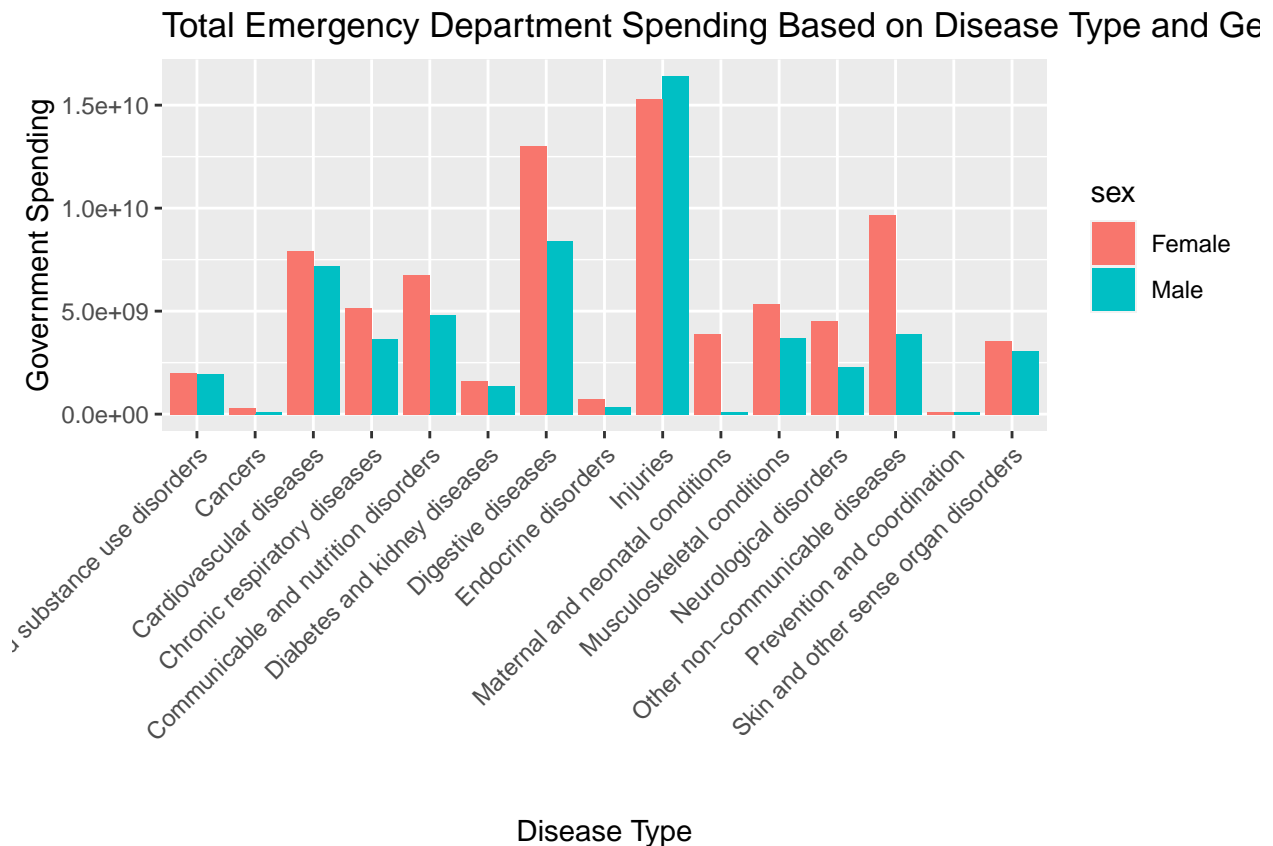
Linear regression model for gender and government spending model

```
spending_malefemale_fit <- linear_reg() %>%
  set_engine("lm") %>%
  fit(mean_all~sex, data = spending_malefemale) %>%
  tidy()
```

Correlation for males vs females given disease type:

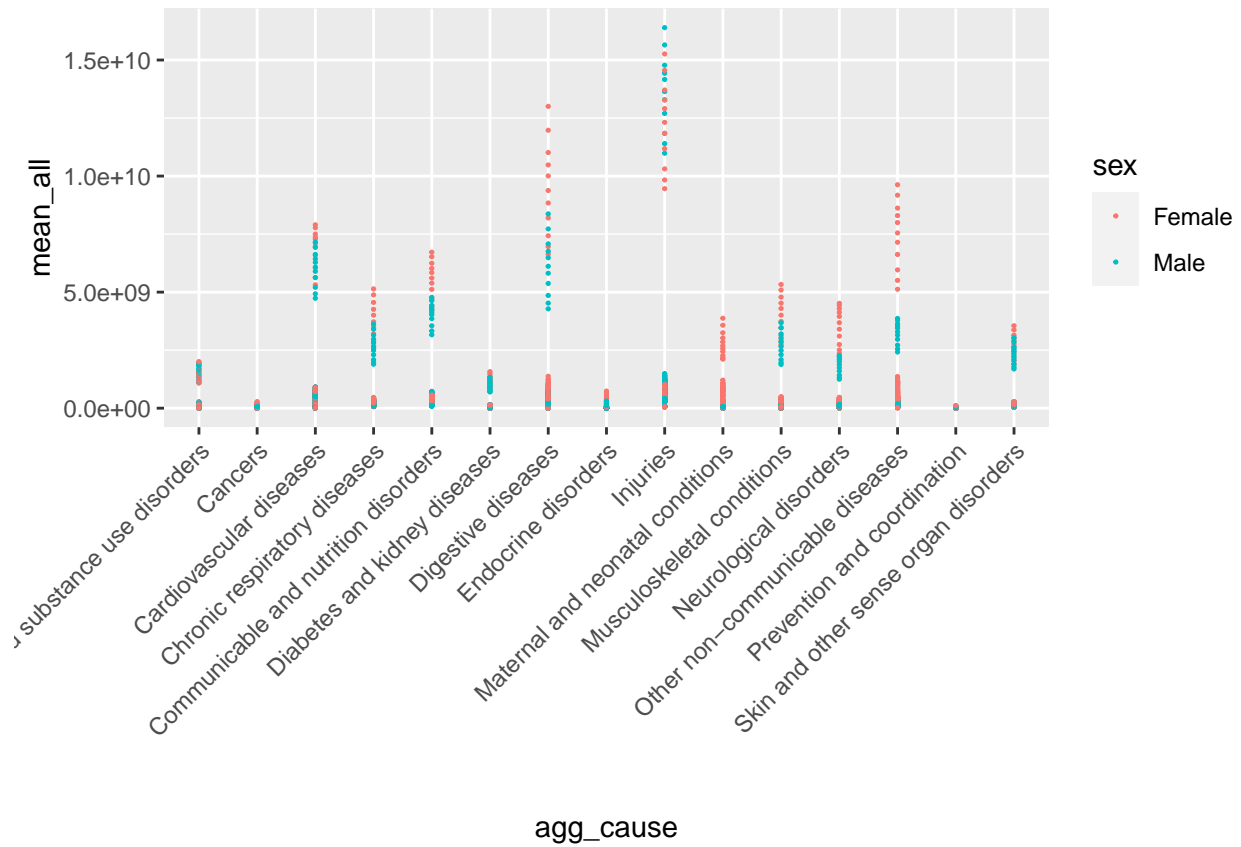
```
spending_malefemale_disease_fit <- linear_reg() %>%
  set_engine("lm") %>%
  fit(mean_all~sex*agg_cause, data = spending_malefemale) %>%
  tidy()
```

```
ggplot(data = spending_malefemale, aes(x = agg_cause, y = mean_all, fill = sex)) +
  geom_bar(position = "dodge", stat = "identity") +
  theme(axis.text.x = element_text(angle = 45,hjust=1)) +
  labs(x = "Disease Type",
       y = "Government Spending",
       title = "Total Emergency Department Spending Based on Disease Type and Gender")
```



!! I was going to fit the linear reg line to this but since the x is categorical it wouldn't make much sense? That's why i added the barchart instead/in addition

```
ggplot(data = spending_malefemale, aes(x = agg_cause, y = mean_all, color = sex)) +
  geom_point(size = 0.25) +
  theme(axis.text.x = element_text(angle = 45,hjust=1))
```



!! Could we do anova for disease type??