An Investigation of Factors Influencing Emergency Healthcare Expenditures

due November 16, 2021 by 11:59 PM

Maggie Lundberg, Riya Mohan, and Izzy Kjaerulff

11/16/2021

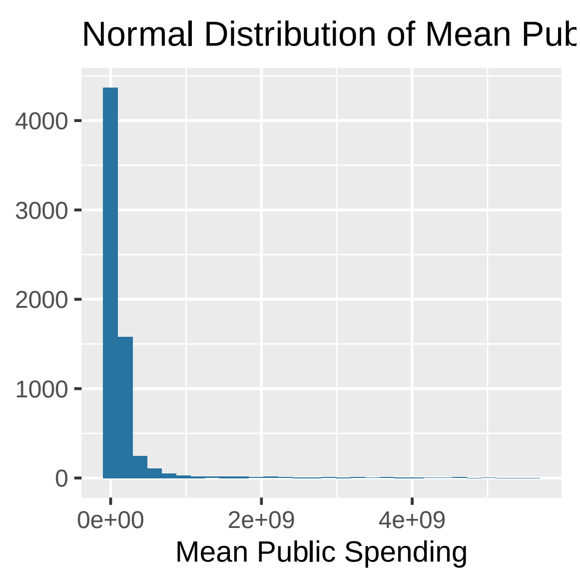
Abstract

Nature of the Data

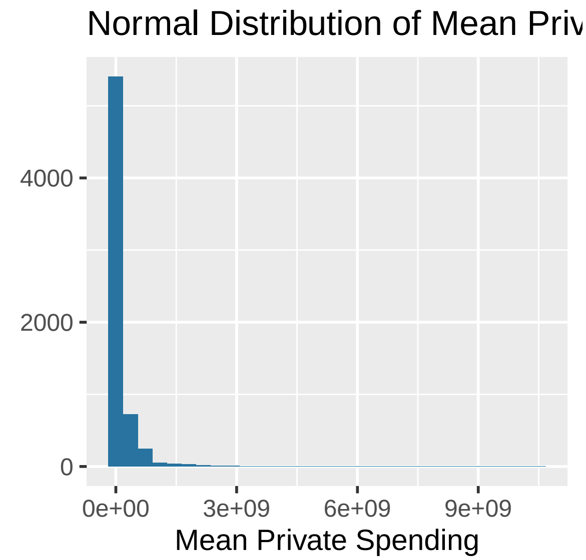
The data includes summary of both male and female spending as "both", so in order to perform an analysis on this data, we decided to exclude the both data points to avoid double counting? It is important to acknowledge that this data only included those who identified as either male or female, so this is not a complete representationof the population.

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

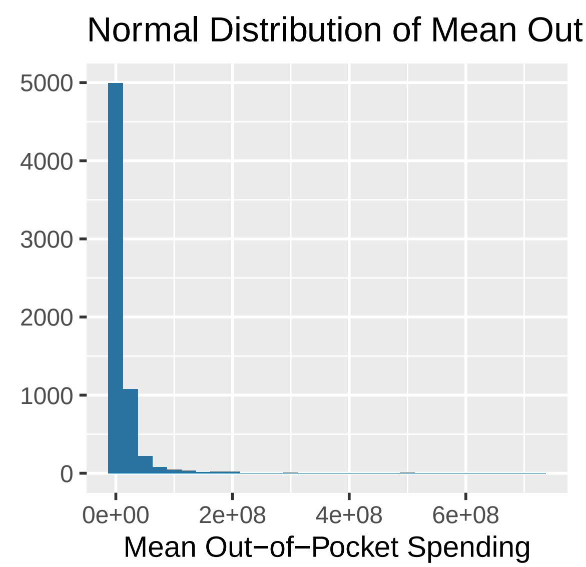
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```




```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

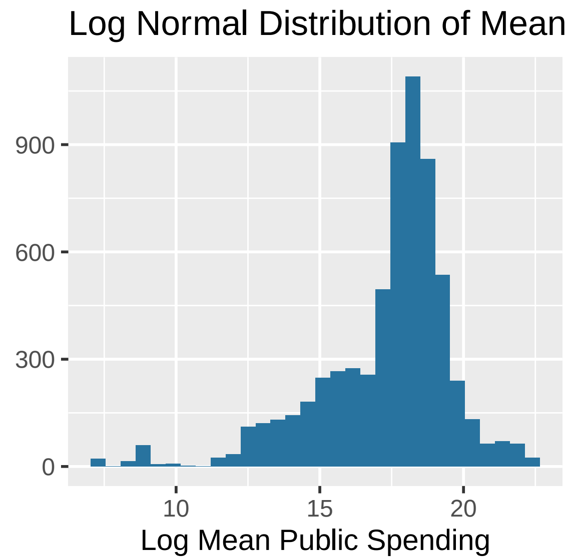



```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

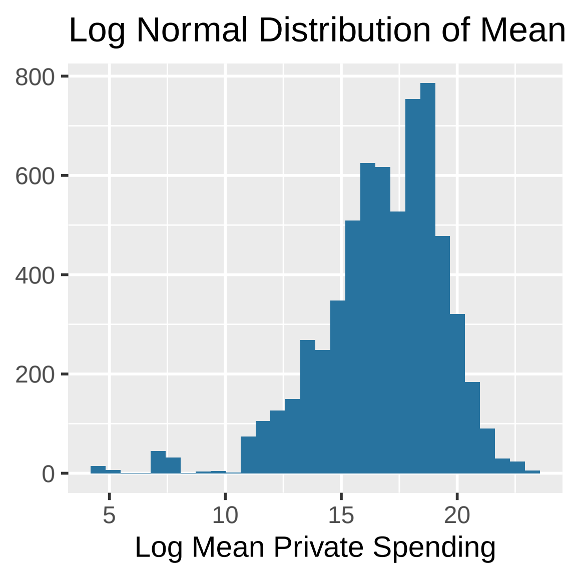


The normal distribution for public spending, private spending, and out-of-pocket pending all show a severe right skew in the data. Therefore, all three variables do not meet the normal distribution assumption needed for many tests, such as ANOVA; however, this can easily be resolved by applying a log transformation to the data to give a fairly normal distribution of the data.

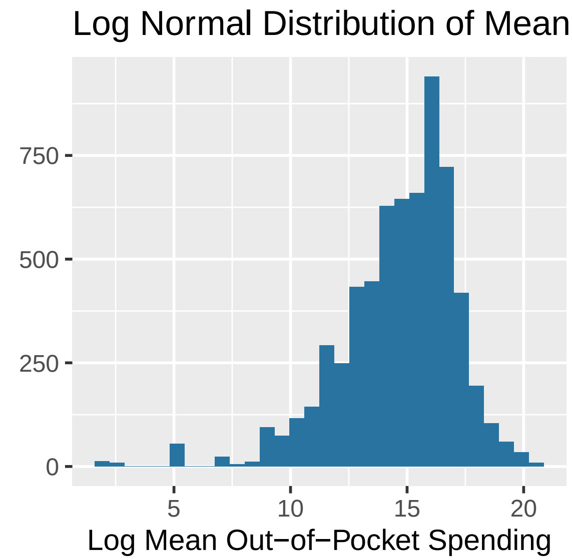
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



`stat_bin()` using `bins = 30`. Pick better value with `binwidth`.



`stat_bin()` using `bins = 30`. Pick better value with `binwidth`.



These graphs of the log distribution of the various spending means appear to be fairly normal in distribution, which means they meet the requirements to be used in various analyses. In order to convert to log scale, those with mean_all, mean_pub, mean_pri, and mean_oop equal to zero must be excluded.

```
spending_malefemale <- spending_malefemale %>%
  filter(mean_all != 0) %>%
  filter(mean_pub != 0) %>%
  filter(mean_pri != 0) %>%
  filter(mean_oop != 0) %>%
  mutate(lmean_all = log(mean_all)) %>%
  mutate(lmean_pub = log(mean_pub)) %>%
  mutate(lmean_pri = log(mean_pri)) %>%
  mutate(lmean_oop = log(mean_oop))
```

Gender

Our first question in this analysis is if males and females spend a different amount of money on emergency services.

First this t-test looks at overall differences in log mean emergency department spending between males and females

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

```
##
##  Welch Two Sample t-test
##
## data:  spending_malefemale$lmean_all by spending_malefemale$sex
## t = 1.4247, df = 6219.5, p-value = 0.1543
## alternative hypothesis: true difference in means between group Female and group Male is not equal to
## 95 percent confidence interval:
##  -0.0315862  0.1996079
## sample estimates:
## mean in group Female   mean in group Male
##             18.06275             17.97874
```

This t-test shows that for mean spending of all emergency services payment types, the p value of 0.1543 (95%

CI -0.0315862, 0.1996079) indicates there is not a significant difference between male and female spending.

Next, we perform a t-test on each type of insurance to see if there is a difference in spending between males and females:

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

```
##
##  Welch Two Sample t-test
##
## data:  spending_malefemale$lmean_pub by spending_malefemale$sex
## t = 1.8142, df = 6201, p-value = 0.0697
## alternative hypothesis: true difference in means between group Female and group Male is not equal to
## 95 percent confidence interval:
##  -0.00833746  0.21532602
## sample estimates:
## mean in group Female   mean in group Male
##             17.40512             17.30162
```

The t-test on emergency services spending for people who have public insurance indicates there is not a significant difference between male and female spending, with p value of 0.0697 (95% CI -0.00833746, 0.21532602).

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

```
##
##  Welch Two Sample t-test
##
## data:  spending_malefemale$lmean_pri by spending_malefemale$sex
## t = 0.70583, df = 6254.9, p-value = 0.4803
## alternative hypothesis: true difference in means between group Female and group Male is not equal to
## 95 percent confidence interval:
##  -0.08283085  0.17603825
## sample estimates:
## mean in group Female   mean in group Male
##             16.82891             16.78231
```

The t-test on emergency services spending for people who have private insurance indicates there is not a significant difference between male and female spending, with p value of 0.4803 (95% CI -0.08283085, 0.17603825).

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

```
##
##  Welch Two Sample t-test
##
## data:  spending_malefemale$lmean_oop by spending_malefemale$sex
## t = 0.9799, df = 6230.6, p-value = 0.3272
## alternative hypothesis: true difference in means between group Female and group Male is not equal to
## 95 percent confidence interval:
##  -0.0615859  0.1846904
## sample estimates:
## mean in group Female   mean in group Male
##             14.66032             14.59877
```

The t-test on emergency services spending for people who pay out of pocket indicates there is not a significant difference between male and female spending, with p value of 0.3272 (95% CI -0.0615859, 0.1846904).

The t-tests for each type of insurance indicate that there is not enough evidence to reject the null hypothesis that emergency department spending is the same for males and females who have public insurance, private insurance, or pay out of pocket, leading us to the conclusion that gender does not influence emergency spending in the forms of payment studied here.

Disease category and Emergency Spending

In order to determine emergency department spending based on disease type, an ANOVA test is performed due to the data for spending on the log scale being normally distributed, relatively similar variance, and independent.

The null hypothesis for this ANOVA test is that the overall mean of spending are the same for each disease category

```
summary(aov(lmean_all~agg_cause,data=spending_malefemale))
```

```
##                Df Sum Sq Mean Sq F value Pr(>F)
## agg_cause      14  19152  1368.0   521.9 <2e-16 ***
## Residuals    6365  16685     2.6
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Based on the p-value here of <2e-16, these data or more extreme data it is highly unlikely the null hypothesis is true. Therefore, we perform step-down tests using a Holm correction for multiple comparisons.

```
diseasepair <- pairwise.t.test(spending_malefemale$lmean_all,    spending_malefemale$agg_cause, p.adj =
sigpairs <- broom::tidy(diseasepair) %>%
  filter(p.value<0.05) %>%
  arrange(group1,group2)
nrow(sigpairs)
```

```
## [1] 92
```

The step-down t tests indicate 92 disease category pairs are different out of 105, indicating most disease categories do differ in the amount of government spending by the emergency department. !!not sure how to interpet anova i dont think this is right

```
meanpubdiseasecatfit <- linear_reg() %>%
  set_engine("lm") %>%
  fit(mean_pub ~ agg_cause, data = spending_malefemale)
  tidy(meanpubdiseasecatfit)
```

```
## # A tibble: 15 x 5
##    term                                    estimate std.error statistic  p.value
##    <chr>                                      <dbl>     <dbl>     <dbl>    <dbl>
##  1 (Intercept)                               8.80e7 22716720.     3.87  1.09e- 4
##  2 agg_causeCancers                         -7.96e7 32126294.    -2.48  1.32e- 2
##  3 agg_causeCardiovascular diseases          3.16e8 32126294.     9.83  1.27e-22
##  4 agg_causeChronic respiratory diseases     1.05e8 32546263.     3.22  1.30e- 3
##  5 agg_causeCommunicable and nutrition d~    1.81e8 32126294.     5.65  1.70e- 8
##  6 agg_causeDiabetes and kidney diseases    -6.83e6 32126294.    -0.212 8.32e- 1
##  7 agg_causeDigestive diseases               2.48e8 32126294.     7.73  1.29e-14
##  8 agg_causeEndocrine disorders             -6.85e7 32126294.    -2.13  3.29e- 2
##  9 agg_causeInjuries                         3.27e8 32126294.    10.2   3.36e-24
## 10 agg_causeMaternal and neonatal condit~    7.91e7 38135560.     2.07  3.82e- 2
```

```
## 11 agg_causeMusculoskeletal conditions       8.03e7 32126294.     2.50  1.25e- 2
## 12 agg_causeNeurological disorders           4.83e7 32126294.     1.50  1.33e- 1
## 13 agg_causeOther non-communicable disea~    1.31e8 32126294.     4.08  4.61e- 5
## 14 agg_causePrevention and coordination     -8.44e7 32126294.    -2.63  8.64e- 3
## 15 agg_causeSkin and other sense organ d~    3.39e7 32126294.     1.06  2.91e- 1
```

Age

!! had to take out the observations with "All Ages" because I think it will just mess up the pairs but let me know what you think or whether you think there's anything we can do with that group

```
spending_noall <- spending_malefemale %>%
  filter(age_group_name != "All Ages")
```

We wonder whether there is a correlation between government healthcare expenditures in the emergency department and age. The age variable is categorical, split into 19 groups that generally include 5 years each, apart from the first (<1 year) and last (85 plus) groups.

To address this question, we began by using an overall test with ANOVA.

Below is an overall test of the null hypothesis that all of the means for age groups across the years are equal, as opposed to the alternative that at least one mean is different.

```
summary(aov(mean_all~age_group_name,data = spending_noall))
```

```
##                   Df    Sum Sq   Mean Sq F value Pr(>F)
## age_group_name    18 2.843e+19 1.579e+18   29.45 <2e-16 ***
## Residuals       6031 3.235e+20 5.364e+16
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

In this F-test (ndf = 18, ddf = 6229), a significant difference among age groups was identified. Our p-value tells us that this data (or data more extreme) would be very unlikely if the null hypothesis were true because it shows statistical significance at an alpha well below 0.05. Therefore, we reject the null hypothesis that the mean expenditures for all age groups are equal.

To see which specific means may be different from one another, we used planned step-down tests with a Holm correction to minimize Type I errors.

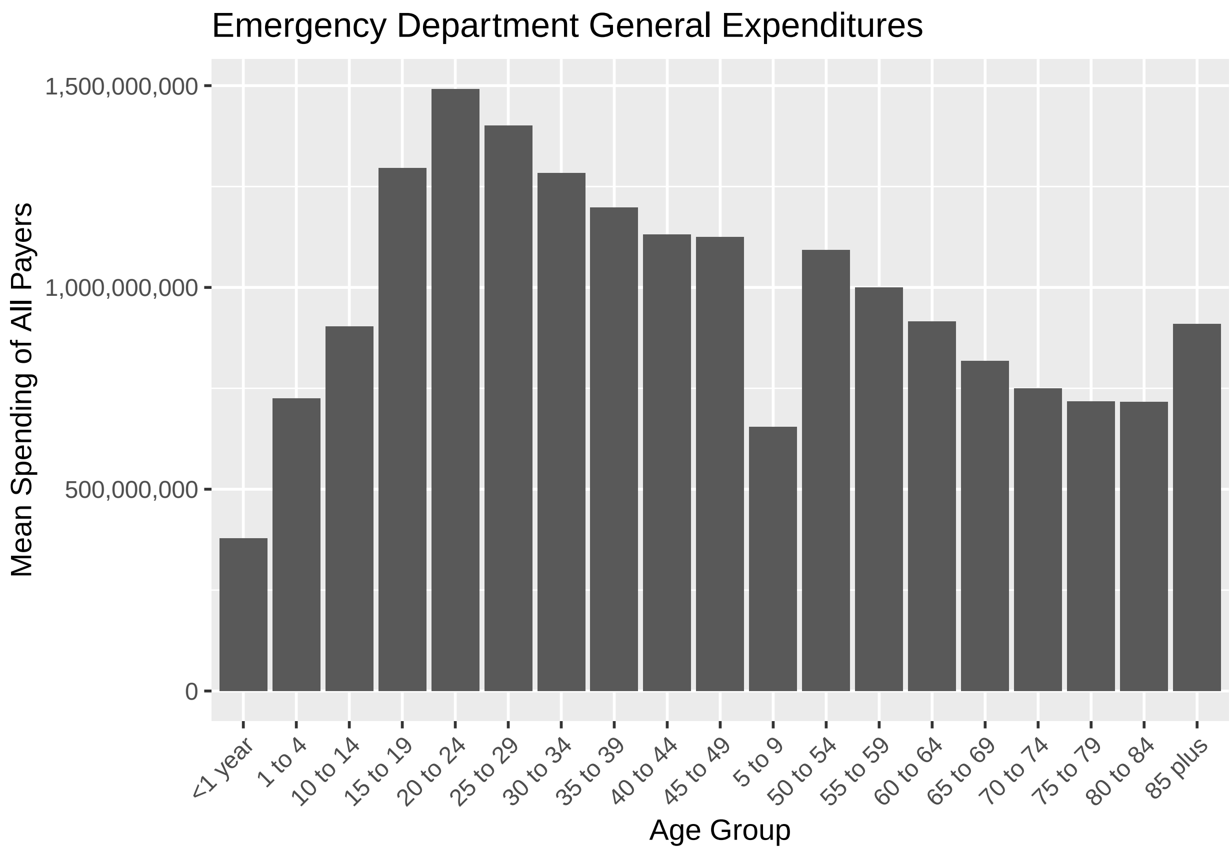
```
agepair <- pairwise.t.test(spending_noall$mean_all, spending_noall$age_group_name, p.adj = "holm")
sigagepairs <- broom::tidy(agepair) %>%
  filter(p.value<0.05) %>%
  arrange(group1,group2)
nrow(sigagepairs)
```

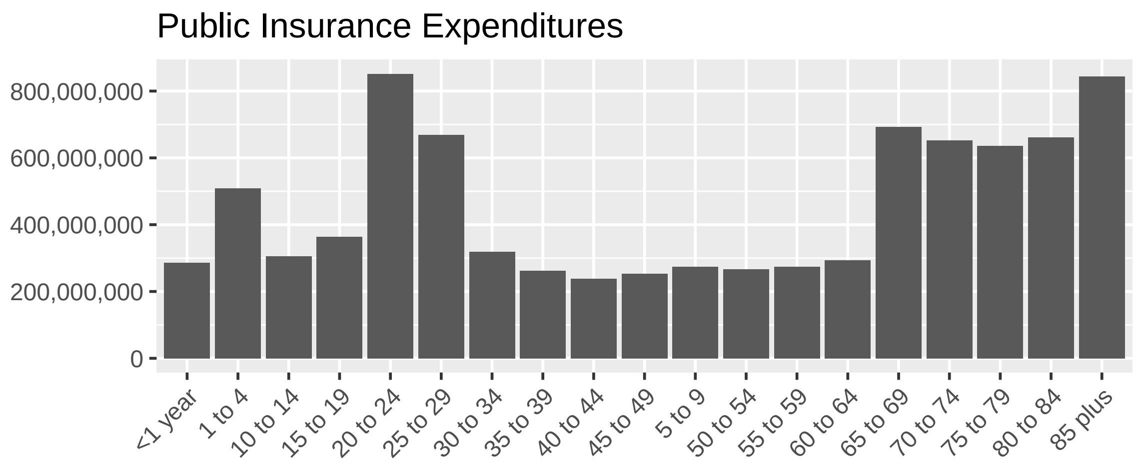
```
## [1] 98
```

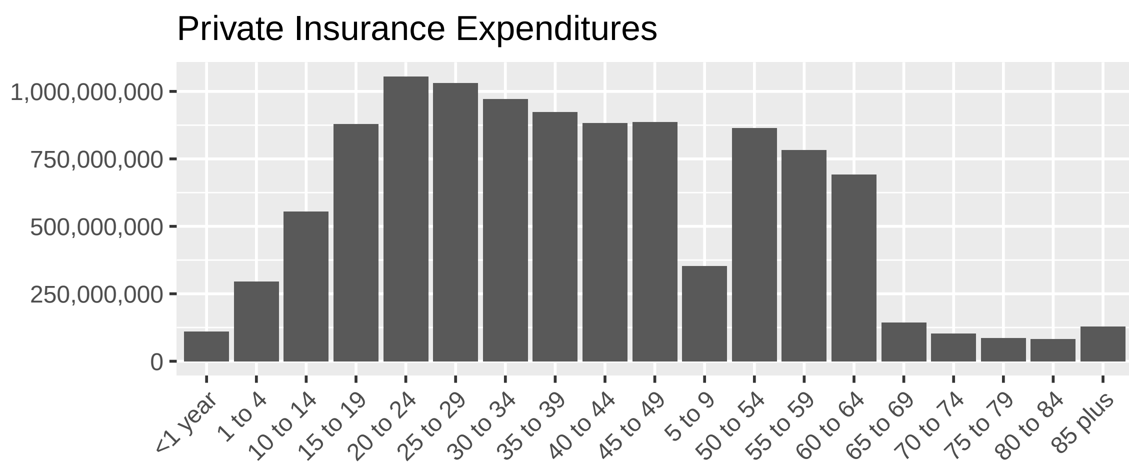
The pairwise t-tests used for our ANOVA step-down tests suggest that there are 97 different age pairs out of the 171 possible combinations. This tells us that more age pairs are different than are similar and that therefore the majority of age group pairs differ in terms of mean expenditures.

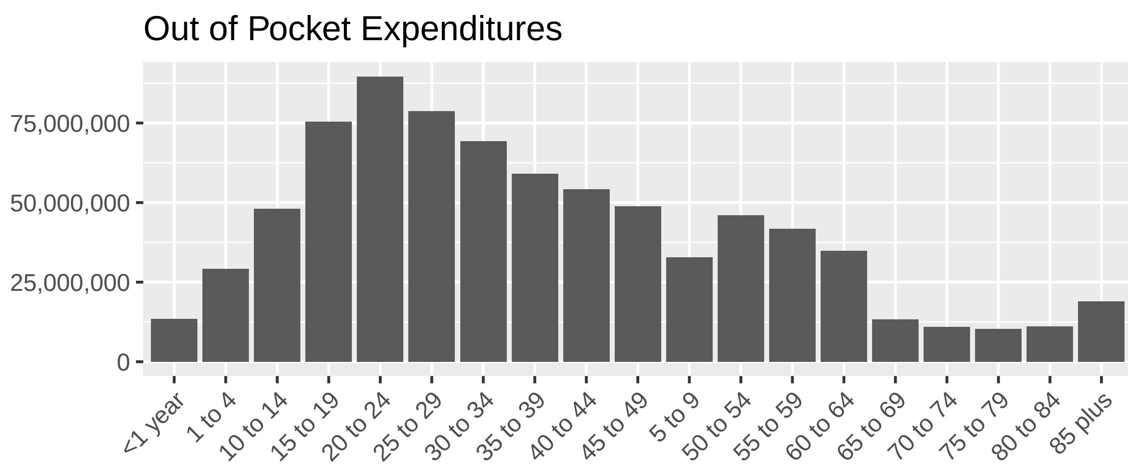
```
  # select the variables want, including the mean for the groups, age_group_name
  # pivot_longer -> cols, names_to = "whateveryouwant", values_to = "customname %>%
  # ggplot(aes(x = age_group_name, y = customname, color = whateveryouwant))

ggplot(data = spending_noall, aes(x = age_group_name, y = mean_all)) +
  geom_bar(position = "dodge", stat = "identity") +
  theme(axis.text.x = element_text(angle = 45,hjust = 1)) +
  scale_y_continuous(labels = scales::comma) +
```

```
  labs(
    x = "Age Group",
    y = "Mean Spending of All Payers",
    title = "Emergency Department General Expenditures"
      )
```









Gender and Age Interaction

```
mainefpub_fit <- linear_reg() %>%
  set_engine("lm") %>%
  fit(mean_pub ~ sex + age_group_id, data = spending)
tidy(mainefpub_fit)
```

```
## # A tibble: 4 x 5
##   term            estimate std.error statistic   p.value
##   <chr>              <dbl>     <dbl>     <dbl>     <dbl>
## 1 (Intercept)   323160777. 12674672.     25.5  6.36e-139
## 2 sexFemale    -137203679. 16690186.     -8.22 2.28e- 16
## 3 sexMale      -203294134. 16690186.    -12.2  6.91e- 34
## 4 age_group_id     788767.   210299.      3.75 1.77e-  4
```

```
glance(mainefpub_fit)$adj.r.squared
```

```
## [1] 0.01649634
```

```
interpub_fit <- linear_reg() %>%
  set_engine("lm") %>%
  fit(mean_pub ~ sex + age_group_id + sex*age_group_id, data = spending)
tidy(interpub_fit)
```

```
## # A tibble: 6 x 5
##   term                    estimate std.error statistic   p.value
##   <chr>                      <dbl>     <dbl>     <dbl>     <dbl>
## 1 (Intercept)             314492259. 14260065.    22.1   2.98e-105
## 2 sexFemale              -131585903. 20166778.    -6.52  7.14e- 11
## 3 sexMale                -182906355. 20166778.    -9.07  1.42e- 19
## 4 age_group_id              1183150.   364221.     3.25  1.16e-  3
## 5 sexFemale:age_group_id    -255587.   515086.    -0.496 6.20e-  1
## 6 sexMale:age_group_id      -927563.   515086.    -1.80  7.18e-  2
```

```
glance(interpub_fit)$adj.r.squared
```

```
## [1] 0.01664197
```

```
mainefpri_fit <- linear_reg() %>%
  set_engine("lm") %>%
  fit(mean_pri ~ sex + age_group_id, data = spending)
tidy(mainefpri_fit)
```

```
## # A tibble: 4 x 5
##   term           estimate std.error statistic   p.value
##   <chr>             <dbl>     <dbl>     <dbl>     <dbl>
## 1 (Intercept)   365176029. 16146361.     22.6  1.86e-110
## 2 sexFemale    -152650378. 21261756.     -7.18 7.50e- 13
## 3 sexMale      -192702232. 21261756.     -9.06 1.51e- 19
## 4 age_group_id    -901887.   267901.     -3.37 7.64e-  4
```

```
glance(mainefpri_fit)$adj.r.squared
```

```
## [1] 0.01001681
```

```
interpri_fit <- linear_reg() %>%
  set_engine("lm") %>%
  fit(mean_pri ~ sex + age_group_id + sex*age_group_id, data = spending)
tidy(interpri_fit)
```

```
## # A tibble: 6 x 5
##   term                    estimate std.error statistic  p.value
##   <chr>                      <dbl>     <dbl>     <dbl>    <dbl>
## 1 (Intercept)             375087738. 18167869.    20.6   9.60e-93
## 2 sexFemale              -166476777. 25693247.    -6.48  9.65e-11
## 3 sexMale                -208610961. 25693247.    -8.12  5.25e-16
## 4 age_group_id             -1352831.   464032.    -2.92  3.56e- 3
## 5 sexFemale:age_group_id     629046.   656240.     0.959 3.38e- 1
## 6 sexMale:age_group_id       723784.   656240.     1.10  2.70e- 1
```

```
glance(interpri_fit)$adj.r.squared
```

```
## [1] 0.009960334
```

```
mainefoop_fit <- linear_reg() %>%
  set_engine("lm") %>%
  fit(mean_oop ~ sex + age_group_id, data = spending)
tidy(mainefoop_fit)
```

```
## # A tibble: 4 x 5
##   term          estimate std.error statistic   p.value
##   <chr>            <dbl>     <dbl>     <dbl>     <dbl>
```

```
## 1 (Intercept)   32646547.  1310040.     24.9  5.50e-133
## 2 sexFemale    -13391357.  1725079.     -7.76 9.14e- 15
## 3 sexMale      -17614550.  1725079.    -10.2  2.34e- 24
## 4 age_group_id    -74643.    21736.     -3.43 5.97e-  4
```

```
glance(mainefoop_fit)$adj.r.squared
```

```
## [1] 0.01226108
```

```
interoop_fit <- linear_reg() %>%
  set_engine("lm") %>%
  fit(mean_oop ~ sex + age_group_id + sex*age_group_id, data = spending)
tidy(interoop_fit)
```

```
## # A tibble: 6 x 5
##   term                    estimate std.error statistic   p.value
##   <chr>                      <dbl>     <dbl>     <dbl>     <dbl>
## 1 (Intercept)            33466867.  1474051.    22.7   2.82e-111
## 2 sexFemale             -14545204.  2084624.    -6.98  3.20e- 12
## 3 sexMale               -18921663.  2084624.    -9.08  1.33e- 19
## 4 age_group_id             -111964.    37649.    -2.97  2.95e-  3
## 5 sexFemale:age_group_id     52495.    53244.     0.986 3.24e-  1
## 6 sexMale:age_group_id       59468.    53244.     1.12  2.64e-  1
```

```
glance(interoop_fit)$adj.r.squared
```

```
## [1] 0.01221011
```

In order to test the possibility that there is a joint interaction of gender and age, a main effects and interaction effects linear regression model has been fit to the data. As a whole, it shows that the interaction of gender and age slightly increases the accuracy of the regression for public and private spending as seen by the increased adjusted R^2 value. However, for out-of-pocket spending, it decreases the adjusted R^2 value. Nevertheless, overall, the adjusted R^2 values for all three types of spending are incredibly low, which further point to our conclusion that age may not affect the level of spending from different sources.

Age and Disease Type Interaction

```
agedismainpub_fit <- linear_reg() %>%
  set_engine("lm") %>%
  fit(mean_pub ~ agg_cause + age_group_id, data = spending)
tidy(agedismainpub_fit)
```

```
## # A tibble: 16 x 5
##    term                                  estimate std.error statistic  p.value
##    <chr>                                    <dbl>     <dbl>     <dbl>    <dbl>
##  1 (Intercept)                             9.99e7 26076508.     3.83  1.28e- 4
##  2 agg_causeCancers                       -1.06e8 36321927.    -2.92  3.48e- 3
##  3 agg_causeCardiovascular diseases        4.21e8 36321927.    11.6   7.57e-31
##  4 agg_causeChronic respiratory diseases   1.40e8 36796801.     3.80  1.45e- 4
##  5 agg_causeCommunicable and nutrition ~   2.42e8 36321927.     6.66  2.89e-11
##  6 agg_causeDiabetes and kidney diseases  -9.10e6 36321927.    -0.251 8.02e- 1
##  7 agg_causeDigestive diseases             3.31e8 36321927.     9.11  9.79e-20
##  8 agg_causeEndocrine disorders           -9.14e7 36321927.    -2.52  1.19e- 2
##  9 agg_causeInjuries                       4.36e8 36321927.    12.0   4.90e-33
## 10 agg_causeMaternal and neonatal condi~   5.21e6 36321927.     0.143 8.86e- 1
## 11 agg_causeMusculoskeletal conditions     1.07e8 36321927.     2.95  3.21e- 3
## 12 agg_causeNeurological disorders         6.44e7 36321927.     1.77  7.61e- 2
```

```
## 13 agg_causeOther non-communicable dise~     1.75e8 36321927.     4.81  1.54e- 6
## 14 agg_causePrevention and coordination     -1.13e8 36321927.    -3.10  1.96e- 3
## 15 agg_causeSkin and other sense organ ~     4.52e7 36321927.     1.25  2.13e- 1
## 16 age_group_id                              7.89e5   205015.     3.85  1.19e- 4
```

```
glance(agedismainpub_fit)$adj.r.squared
```

```
## [1] 0.06530174
```

```
agedisinterpub_fit <- linear_reg() %>%
  set_engine("lm") %>%
  fit(mean_pub ~ agg_cause + age_group_id + agg_cause*age_group_id, data = spending)
tidy(agedisinterpub_fit)
```

```
## # A tibble: 30 x 5
##    term                                     estimate std.error statistic  p.value
##    <chr>                                       <dbl>     <dbl>     <dbl>    <dbl>
##  1 (Intercept)                                1.24e8 31019162.     4.01  6.16e- 5
##  2 agg_causeCancers                          -1.15e8 43867720.    -2.62  8.81e- 3
##  3 agg_causeCardiovascular diseases           3.11e8 43867720.     7.08  1.55e-12
##  4 agg_causeChronic respiratory diseases      1.22e8 44168605.     2.76  5.78e- 3
##  5 agg_causeCommunicable and nutrition ~      2.29e8 43867720.     5.21  1.93e- 7
##  6 agg_causeDiabetes and kidney diseases     -2.30e7 43867720.    -0.524 6.01e- 1
##  7 agg_causeDigestive diseases                3.07e8 43867720.     7.01  2.57e-12
##  8 agg_causeEndocrine disorders              -1.00e8 43867720.    -2.28  2.25e- 2
##  9 agg_causeInjuries                          3.72e8 43867720.     8.49  2.34e-17
## 10 agg_causeMaternal and neonatal condi~      1.38e7 43867720.     0.314 7.54e- 1
## # ... with 20 more rows
```

```
glance(agedisinterpub_fit)$adj.r.squared
```

```
## [1] 0.06778399
```

```
agedismainpri_fit <- linear_reg() %>%
  set_engine("lm") %>%
  fit(mean_pri ~ agg_cause + age_group_id, data = spending)
tidy(agedismainpri_fit)
```

```
## # A tibble: 16 x 5
##    term                                   estimate std.error statistic   p.value
##    <chr>                                     <dbl>     <dbl>     <dbl>     <dbl>
##  1 (Intercept)                              9.57e7 32514561.     2.94  3.26e-  3
##  2 agg_causeCancers                        -6.93e7 45289481.    -1.53  1.26e-  1
##  3 agg_causeCardiovascular diseases         2.20e8 45289481.     4.86  1.20e-  6
##  4 agg_causeChronic respiratory diseas~     1.02e8 45881596.     2.22  2.65e-  2
##  5 agg_causeCommunicable and nutrition~     1.72e8 45289481.     3.80  1.44e-  4
##  6 agg_causeDiabetes and kidney diseas~    -4.50e7 45289481.    -0.993 3.20e-  1
##  7 agg_causeDigestive diseases              4.68e8 45289481.    10.3   7.16e- 25
##  8 agg_causeEndocrine disorders            -5.93e7 45289481.    -1.31  1.91e-  1
##  9 agg_causeInjuries                        1.01e9 45289481.    22.3   1.06e-107
## 10 agg_causeMaternal and neonatal cond~    -1.36e7 45289481.    -0.299 7.65e-  1
## 11 agg_causeMusculoskeletal conditions      1.27e8 45289481.     2.81  4.95e-  3
## 12 agg_causeNeurological disorders          8.64e7 45289481.     1.91  5.66e-  2
## 13 agg_causeOther non-communicable dis~     3.02e8 45289481.     6.68  2.56e- 11
## 14 agg_causePrevention and coordination    -7.12e7 45289481.    -1.57  1.16e-  1
## 15 agg_causeSkin and other sense organ~     8.28e7 45289481.     1.83  6.75e-  2
## 16 age_group_id                            -9.03e5   255631.    -3.53  4.14e-  4
```

```r
glance(agedismainpri_fit)$adj.r.squared
```

```
## [1] 0.09862747
```

```r
agedisinterpri_fit <- linear_reg() %>%
  set_engine("lm") %>%
  fit(mean_pri ~ agg_cause + age_group_id + agg_cause*age_group_id, data = spending)
tidy(agedisinterpri_fit)
```

```
## # A tibble: 30 x 5
##    term                                   estimate std.error statistic  p.value
##    <chr>                                     <dbl>     <dbl>     <dbl>    <dbl>
##  1 (Intercept)                              8.36e7 38715024.     2.16  3.08e- 2
##  2 agg_causeCancers                        -7.68e7 54751312.    -1.40  1.61e- 1
##  3 agg_causeCardiovascular diseases         2.24e8 54751312.     4.08  4.46e- 5
##  4 agg_causeChronic respiratory diseases    1.08e8 55126847.     1.96  5.01e- 2
##  5 agg_causeCommunicable and nutrition ~    1.87e8 54751312.     3.41  6.49e- 4
##  6 agg_causeDiabetes and kidney diseases   -5.10e7 54751312.    -0.932 3.51e- 1
##  7 agg_causeDigestive diseases              5.13e8 54751312.     9.37  8.87e-21
##  8 agg_causeEndocrine disorders            -6.58e7 54751312.    -1.20  2.29e- 1
##  9 agg_causeInjuries                        1.11e9 54751312.    20.2   8.41e-89
## 10 agg_causeMaternal and neonatal condi~   -1.43e7 54751312.    -0.260 7.95e- 1
## # ... with 20 more rows
```

```r
glance(agedisinterpri_fit)$adj.r.squared
```

```
## [1] 0.09927656
```

```r
agedismainoop_fit <- linear_reg() %>%
  set_engine("lm") %>%
  fit(mean_oop ~ agg_cause + age_group_id, data = spending)
tidy(agedismainoop_fit)
```

```
## # A tibble: 16 x 5
##    term                                   estimate std.error statistic  p.value
##    <chr>                                     <dbl>     <dbl>     <dbl>    <dbl>
##  1 (Intercept)                              1.42e7  2659045.     5.32  1.04e- 7
##  2 agg_causeCancers                        -1.20e7  3703779.    -3.25  1.17e- 3
##  3 agg_causeCardiovascular diseases         1.03e7  3703779.     2.77  5.61e- 3
##  4 agg_causeChronic respiratory diseases    3.71e6  3752202.     0.989 3.23e- 1
##  5 agg_causeCommunicable and nutrition d~   1.81e7  3703779.     4.90  9.79e- 7
##  6 agg_causeDiabetes and kidney diseases   -8.09e6  3703779.    -2.18  2.89e- 2
##  7 agg_causeDigestive diseases              3.40e7  3703779.     9.17  5.55e-20
##  8 agg_causeEndocrine disorders            -9.97e6  3703779.    -2.69  7.12e- 3
##  9 agg_causeInjuries                        6.76e7  3703779.    18.2   3.80e-73
## 10 agg_causeMaternal and neonatal condit~  -6.02e6  3703779.    -1.63  1.04e- 1
## 11 agg_causeMusculoskeletal conditions      5.98e6  3703779.     1.61  1.06e- 1
## 12 agg_causeNeurological disorders          1.64e6  3703779.     0.442 6.58e- 1
## 13 agg_causeOther non-communicable disea~   2.64e7  3703779.     7.13  1.08e-12
## 14 agg_causePrevention and coordination    -1.22e7  3703779.    -3.28  1.04e- 3
## 15 agg_causeSkin and other sense organ d~   2.73e6  3703779.     0.736 4.62e- 1
## 16 age_group_id                            -7.47e4    20906.    -3.57  3.53e- 4
```

```r
glance(agedismainoop_fit)$adj.r.squared
```

```
## [1] 0.08631926
```

```r
agedisinteroop_fit <- linear_reg() %>%
  set_engine("lm") %>%
  fit(mean_oop ~ agg_cause + age_group_id + agg_cause*age_group_id, data = spending)
tidy(agedisinteroop_fit)
```

```
## # A tibble: 30 x 5
##    term                                     estimate std.error statistic  p.value
##    <chr>                                       <dbl>     <dbl>     <dbl>    <dbl>
##  1 (Intercept)                               1.37e7  3167013.      4.34  1.43e- 5
##  2 agg_causeCancers                         -1.32e7  4478833.     -2.96  3.11e- 3
##  3 agg_causeCardiovascular diseases          9.34e6  4478833.      2.09  3.71e- 2
##  4 agg_causeChronic respiratory diseases     3.66e6  4509553.      0.811 4.17e- 1
##  5 agg_causeCommunicable and nutrition d~    1.97e7  4478833.      4.41  1.06e- 5
##  6 agg_causeDiabetes and kidney diseases    -9.10e6  4478833.     -2.03  4.23e- 2
##  7 agg_causeDigestive diseases               3.70e7  4478833.      8.25  1.74e-16
##  8 agg_causeEndocrine disorders             -1.10e7  4478833.     -2.46  1.38e- 2
##  9 agg_causeInjuries                         7.30e7  4478833.     16.3   6.06e-59
## 10 agg_causeMaternal and neonatal condit~   -6.49e6  4478833.     -1.45  1.47e- 1
## # ... with 20 more rows
```

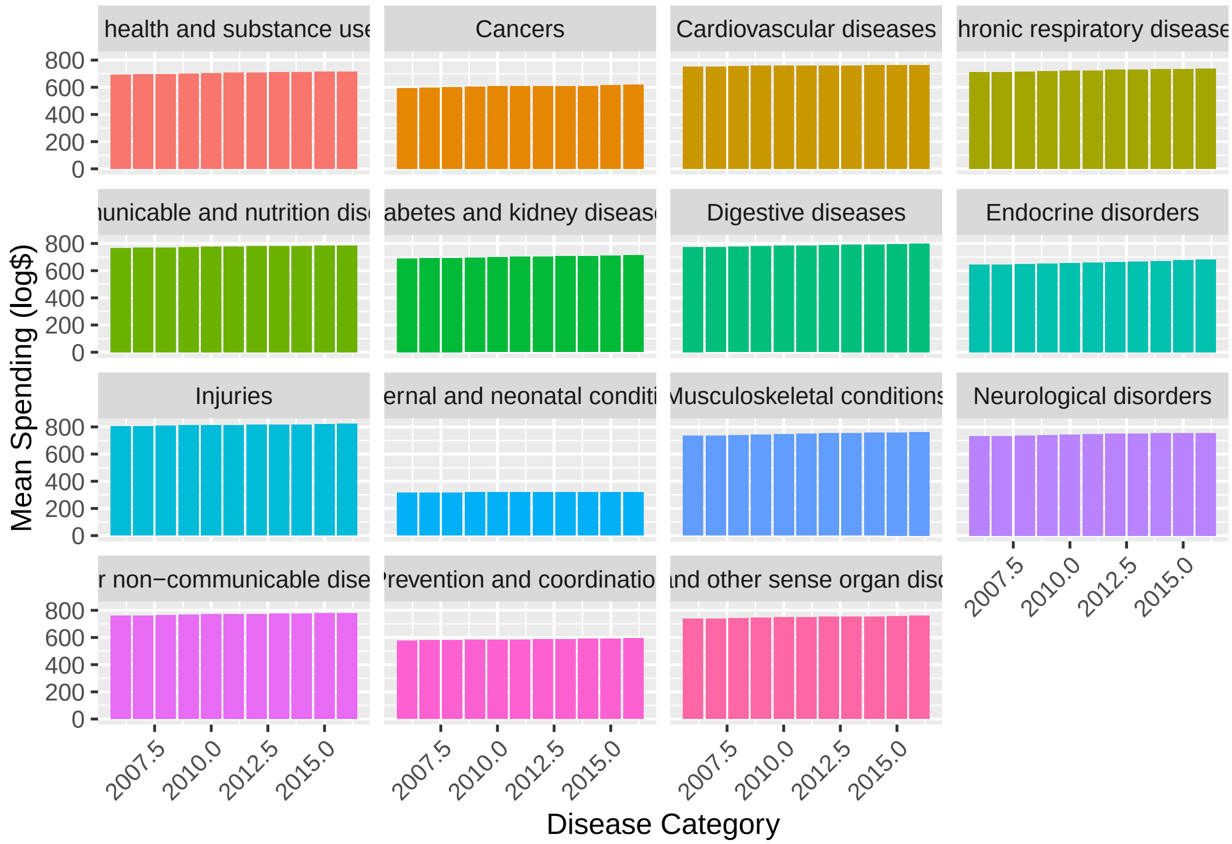
```r
glance(agedisinteroop_fit)$adj.r.squared
```

```
## [1] 0.08646145
```

##Spending Over Time

!! I kinda like this but idk if it adds anything but it is fun, need to make the words smaller so you can read it

```r
spending_malefemale %>%
  ggplot(aes(x = year_id,
             y = lmean_all,
             fill = agg_cause)) +
  geom_bar(stat = "identity") +
  facet_wrap(~agg_cause) +
  theme(axis.text.x = element_text(angle = 45,hjust = 1), legend.position = "none") +
  labs(x = "Disease Category",
       y = "Mean Spending (log$)")
```




```
spendingovertime_fit <- linear_reg() %>%
  set_engine("lm") %>%
  fit(lmean_all ~ year_id, data = spending_malefemale)
tidy(spendingovertime_fit)
```

```
## # A tibble: 2 x 5
##   term        estimate std.error statistic       p.value
##   <chr>          <dbl>     <dbl>     <dbl>         <dbl>
## 1 (Intercept) -92.1      18.8        -4.89 0.00000102
## 2 year_id       0.0548    0.00936     5.85 0.00000000516
```