Midterm Project

Ashley Dickson, Simran Singh, Qi Tan, Jason Theiling

10/30/2020

Wellness Product Market Sizing

Task 1: A First Look at The Data

1) Read in the data from 2014 and check the dimensions of this table.

The 2014 data set contains mortality data, including demographic details, age, and ICD-10 codes (medical diagnosis codes attributable to the cause of death).

Dimensions of 2014 Mortality Data set:

- Number of Rows: 2,631,171
- Number of Columns: 77
- Interpretation: This data set contains records for a total of 2,631,171 individuals who died in the year 2014. Each row represents a unique mortality event or individual. There are 77 columns which includes information such as demographic details.

```
# Read in the 2014 Mortality Data set
full_2014 = read_csv("2014_data.csv")
```

```
## Warning: One or more parsing issues, call `problems()` on your data frame for details,
## e.g.:
##   dat <- vroom(...)
##   problems(dat)
```

```
## Rows: 2631171 Columns: 77
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (45): education_1989_revision, month_of_death, sex, detail_age, age_reco...
## dbl (16): resident_status, education_2003_revision, education_reporting_flag...
## lgl (16): age_substitution_flag, entity_condition_14, entity_condition_15, e...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
# Get the dimensions for the 2014 Mortality Data set
dimensions_2014 = dim(full_2014)
```

2) You decide that this data_frame takes much too long to manipulate and decide to take a subset of it in order to make later steps more manageable. Take a random sample of 100,000 rows from this data_frame.

```
# Take a random sample of 100,000 rows from the 2014 Data set
sample_2014 = full_2014[sample(nrow(full_2014), 100000), seed = 123]
```

3) While this is still a sizable sample, you worry that this subset may not be representative of the full data set. To check this, you want to look at some summary statistics from both tables to ensure they are still representative. Look at the average age and %Female in both tables. In your opinion, is there significant variation?

NOTE: A bit of cleaning is required before looking at average age. Per the data dictionary, *detail_age* shows the individual's age in years only when detail_age_type == 1. Do some quality checks (e.g. sort in descending/ascending order to find extreme values) to make sure this makes sense. Provide rationale for any other exclusions.

In the data cleaning process, several steps were taken to prepare the 2014 data set for analysis. First, we filtered the data to exclude individuals who were not over a year old and those whose ages were not stated. The goal was to focus on individuals whose ages were recorded in years, ensuring that the mean calculation would accurately represent the data without introducing bias. herefore, detailed age types over "1" and age recode labeled as "52" were removed from the data set. Additionally, leading zeros in the "detail_age" column were removed, and the column was converted to a numeric variable to ensure the calculation of the mean.

It is important to note that the same cleaning process was applied to the random sample 2014 data set to maintain consistency in data cleaning across both data sets.

Once the cleaning was completed for both the random sample and the full 2014 data set, the analysis focused on calculating the average age and the percentage of females in both data sets. The results indicated that the mean age did not significantly vary between the sample and the full data set, suggesting that the random sample is a representative subset of the complete data set. This consistency in data cleaning and the comparability of the results between the sample and the full data set validated the use of the random sample for further analysis.

Below we have listed the average age and percentage of females in the original full 2014 data set and the random sample taken from the original full 2014 data set.

Original 2014 Data set -

- Average Age: 73.79 years
- %Female: 49.44%

Random Sample

- Average Age: 73.78 years
- %Female: 49.15%

```
# Filter values in sample 2014 data set to remove individuals who are not over a year old and individua
sample_2014_cleaned = sample_2014 %>%
  filter(detail_age_type == 1) %>%
  filter(age_recode_52 < 52)

# Remove leading zeros from the " detail_age" column
```

```r
sample_2014_cleaned$detail_age = str_replace(sample_2014_cleaned$detail_age, "^0+", "")

# Convert 'detail_age' to numeric variable
sample_2014_cleaned$detail_age = as.numeric(sample_2014_cleaned$detail_age)

# Calculate the average age and %Female for the cleaned random sample
avg_sample_2014 = sample_2014_cleaned %>%
  summarise(sample_mean_age = mean(detail_age, na.rm = TRUE,),
            sample_female_pct = sum(sex == 'F') / n() * 100)

# Filter values in full 2014 data set to remove individuals who are not over a year old and individuals
full_2014_cleaned = full_2014 %>%
  filter(detail_age_type == 1) %>%
  filter(age_recode_52 < 52)

# Remove leading zeros from the " detail_age" column
full_2014_cleaned$detail_age = str_replace(full_2014_cleaned$detail_age, "^0+", "")

# Convert 'detail_age' to numeric variable
full_2014_cleaned$detail_age = as.numeric(full_2014_cleaned$detail_age)

# Calculate the average age and %Female for the cleaned random sample
avg_full_2014 = full_2014_cleaned %>%
  summarise(full_mean_age = mean(detail_age, na.rm = TRUE),
            full_female_pct = sum(sex == 'F', na.rm = TRUE) / n() *100)
```

4) Satisfied with the degree of variability (or lack thereof) in the previous step, you remove the larger data_frame (using rm()) and repeat steps 1-2 with the 2015 data_frame.

The 2015 data set contains mortality data, including demographic details, age, and ICD-10 codes (medical diagnosis codes attributable to the cause of death).

Dimensions of 2015 Mortality Data set:

- Number of Rows: 2,718,198
- Number of Columns: 77
- Interpretation: This data set contains records for a total of 2,718,198 individuals who died in the year 2015. Similar to the 2014 data set, each row represents a unique mortality event or individual. The data set also consists of 77 columns, which are structured in a similar way to the 2014 data set. The 2015 data set is slightly larger in terms of the number of records compared to the 2014 data set.

```r
# Read in the 2015 Mortality Data set
full_2015 = read_csv("2015_data.csv")
```

```
## Warning: One or more parsing issues, call `problems()` on your data frame for details,
## e.g.:
##   dat <- vroom(...)
##   problems(dat)

## Rows: 2718198 Columns: 77
## -- Column specification --------------------------------------------------------
## Delimiter: ","
```

```
## chr (41): education_1989_revision, month_of_death, sex, detail_age, age_reco...
## dbl (16): resident_status, education_2003_revision, education_reporting_flag...
## lgl (20): age_substitution_flag, entity_condition_12, entity_condition_13, e...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

In the data cleaning process, several steps were taken to prepare the 2015 data set for analysis. First, we filtered the data to exclude individuals who were not over a year old and those whose ages were not stated. The goal was to focus on individuals whose ages were recorded in years, ensuring that the mean calculation would accurately represent the data without introducing bias. Therefore, detailed age types over "1" and age recode labeled as "52" were removed from the data set. Additionally, leading zeros in the "detail_age" column were removed, and the column was converted to a numeric variable to ensure the calculation of the mean.

It is important to note that the same cleaning process was applied to the random sample 2015 data set to maintain consistency in data cleaning across both data sets.

Once the cleaning was completed for both the random sample and the full 2015 data set, the analysis focused on calculating the average age and the percentage of females in both data sets. The results indicated that the mean age did not significantly vary between the sample and the full data set, suggesting that the random sample is a representative subset of the complete data set. This consistency in data cleaning and the comparability of the results between the sample and the full data set validated the use of the random sample for further analysis.

Below we have listed the average age and percentage of females in the original full 2015 data set and the random sample taken from the original full 2015 data set.

Original 2015 Data set

- Average Age: 73.84 years
- %Female: 49.38%

Random Sample

- Average Age: 73.88 years
- %Female: 49.29%

```r
# Remove the large 2014 data set
rm(full_2014)

# Take a random sample of 100,000 rows from the 2015 Data set
sample_2015 = full_2015[sample(nrow(full_2015), 100000), seed = 123]

# Filter values in sample 2015 data set to remove individuals who are not over a year old and individua
sample_2015_cleaned = sample_2015 %>%
  filter(detail_age_type == 1) %>%
  filter(age_recode_52 < 52)

# Remove leading zeros from the " detail_age" column
sample_2015_cleaned$detail_age = str_replace(sample_2015_cleaned$detail_age, "^0+", "")

# Convert 'detail_age' to numeric variable
```

```r
sample_2015_cleaned$detail_age = as.numeric(sample_2015_cleaned$detail_age)

# Calculate the average age and %Female for the cleaned random sample
avg_sample_2015 = sample_2015_cleaned %>%
  summarise(mean_sample_age = mean(detail_age, na.rm = TRUE,),
            sample_female_pct = sum(sex == 'F') / n() * 100)

# Filter values in full 2015 data set to remove individuals who are not over a year old and individuals
full_2015_cleaned = full_2015 %>%
  filter(detail_age_type == 1) %>%
  filter(age_recode_52 < 52)

# Remove leading zeros from the " detail_age" column
full_2015_cleaned$detail_age = str_replace(full_2015_cleaned$detail_age, "^0+", "")

# Convert 'detail_age' to numeric variable
full_2015_cleaned$detail_age = as.numeric(full_2015_cleaned$detail_age)

# Calculate the average age and %Female for the cleaned random sample
avg_full_2015 = full_2015_cleaned %>%
  summarise(mean_full_age = mean(detail_age, na.rm = TRUE,),
            full_female_pct = sum(sex == 'F') / n() *100)
```

5) Combine these two data_frames into a single data_frame

```r
# Combine the data frames by stacking the rows
sample_2014_2015_cleaned = rbind(sample_2014_cleaned, sample_2015_cleaned)
```

Task 2: How does age of death vary by Male/Female?

1) You are aware of some basic medical knowledge that women tend to have a longer life expectancy than men, so you generate two more questions:

a) What is the average age of death for women versus men in these data?

Using the combined data set created in Task 1, Step 5, the average age of death for women is 77.12 years, whereas the average age of death for men is 70.65 years. This data indicates that, on average, women tend to live longer than men in the analyzed data set, with a notable difference in the average age at the time of death between the two genders.

```r
#  Using the combined data set, group the genders and calculate the mean of the column within each grou
sample_2014_2015_cleaned_summarize = sample_2014_2015_cleaned %>%
   group_by(sex) %>%
   summarize(mean_age_death=mean(detail_age))
```

b) Does this relationship hold when we look at the average age by year? Note any differences.

When examining the average age of death by year, we can observe that the relationship between the average age of death for women and men holds consistently across both 2014 and 2015. In 2014, the average age of death for women was 77.09 years, while for men, it was 70.60 years. Similarly, in 2015, the average age of death for women was 77.14 years, and for men, it was 70.70 years.

These findings reaffirm the trend observed in the overall data, with women consistently having a higher average age at the time of death compared to men, regardless of the specific year under consideration.

```
#  Using the combined data set, group the genders by year and calculate the mean of the column within e
sample_2014_2015_cleaned_summarize_byyear = sample_2014_2015_cleaned %>%
   group_by(sex, current_data_year) %>%
   summarize(mean_age_death=mean(detail_age))
```

```
## `summarise()` has grouped output by 'sex'. You can override using the `.groups`
## argument.
```

2) Unsatisfied with the simple averages, you would like to see the distribution of age of mortality over this population over both years. Plot the distribution of age for Male and Female and note any differences. What might explain the difference in averages?

2014 Density Plot:

(a) Any differences?

The differences in the distribution of age of mortality between males and females in 2014 are as follows:

1. **Central Tendency (Mean and Median):**
   - Females have a higher mean age (77.09 years) compared to males (70.60 years). This suggests that, on average, females tend to live longer than males.
   - The median for females (81 years) is notably higher than that for males (73 years). This reflects that the middle value of the age distribution is higher for females
2. **Spread and Variability (Standard Deviation and Variance):**
   - Females have a smaller standard deviation (16.05) and variance (257.55) compared to males (standard deviation of 17.33 and variance of 300.50). This indicates that the age distribution for females is less spread out and less variable compared to males.
3. **Minimum and Maximum Ages:**
   - The minimum and maximum ages for both females and males are similar, ranging from 1 to 113 for females and 1 to 110 for males.
4. **Interquartile Range (IQR):**
   - The IQR for females (21) is slightly smaller than that for males (23). This means that the middle 50% of the female age distribution falls within a narrower age range compared to males.
5. **Mode:**
   - The mode (most common age) for females is 87 years, with 1,716 women at this age. For males, the mode is 83 years, with 1,354 men at this age. This suggests that 87 is the most common mortality age for females, while 83 is the most common mortality age for males.
6. **Shape of the Distribution:**
   - The female age distribution is described as unimodal and left-skewed, indicating that most women are clustered towards the older age groups. The left skew implies that there are relatively fewer women who pass at a "younger" age.

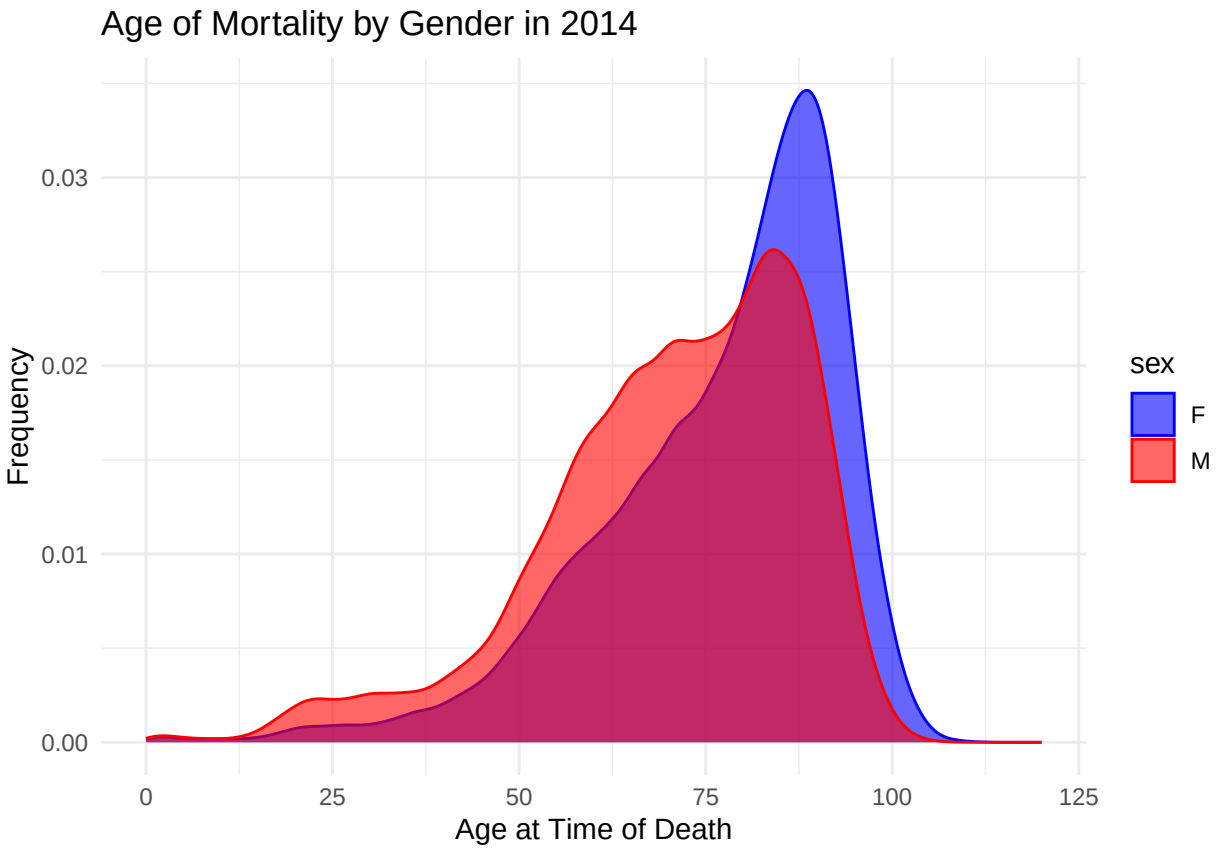
- The male age distribution is slightly bimodal and left-skewed, suggesting the presence of two modes, with most men clustered around the 83-year and younger age groups. The left skew implies that there are relatively fewer older men.

(b) What might explain the difference in averages?

There a few possible explanations for the differences in the averages:

- Biological differences between the sexes, such as genetics and hormones, can influence life expectancy. Women tend to have longer life expectancies, which may explain the higher mean and median age for females.
- Behavioral and health-related factors, such as smoking, diet, and access to healthcare, can contribute to differences in life expectancy between males and females.
- The differences in mode and skewness may be due to variations in the gender gap in mortality at different age groups. Factors like disease susceptibility, accidents, and lifestyle choices can affect this gap.

```r
# Plot the distribution of age of mortality for both genders for a comprehensive view in 2014
sample_2014_cleaned %>%
  ggplot(aes(x = detail_age, color = sex, fill = sex)) +
  geom_density(alpha = 0.6) +
  xlim(0, 120) +
  scale_fill_manual(values = c("blue", "red")) +
  scale_color_manual(values = c("blue", "red")) +
  theme_minimal() +
  labs(title = "Age of Mortality by Gender in 2014", x = "Age", y = "Density") +
  ylab("Frequency") +
  xlab("Age at Time of Death")
```



In order to find the key differences between Females and Males, we found the summary statistics for each density plot.

```
# Find the summary statistics for the Density plot for Sex in 2014
sample_2014_cleaned_summarize_bysex = sample_2014_cleaned %>%
   group_by(sex) %>%
   summarize(mean_age_death=mean(detail_age), median_age_of_death=median(detail_age), sd_age_of_death=sd
             min_age_death=min(detail_age), max_age_of_death=max(detail_age), iqr_age_of_death=IQR(detai
```

2015 Distribution Charts:

(a) Any differences?

The differences in the distribution of age of mortality between males and females in 2015 are as follows:

1. **Central Tendency (Mean and Median):**
   - Females have a higher mean age (77.14 years) compared to males (70.70 years). This supports the consistent finding that, on average, females tend to live longer than males.
   - The median for females (81 years) remains notably higher than that for males (73 years). This points to a higher central age tendency for females, reflecting greater longevity.
2. **Spread and Variability (Standard Deviation and Variance):**
   - Females continue to exhibit a smaller standard deviation (16.22) and variance (263.06) compared to males (standard deviation of 17.33 and variance of 300.28). This consistency indicates that the age distribution for females is less spread out and less variable than that for males.

3. **Minimum and Maximum Ages:**

    - The minimum and maximum ages for both females and males are similar, ranging from 1 to 114 for females and 1 to 111 for males.

4. **Interquartile Range (IQR):**

    - The IQR for females (21 years) remains slightly smaller than that for males (23 years), indicating that the middle 50% of the female age distribution falls within a narrower age range compared to males.

5. **Shape of the Distribution:**

    - The female age distribution remains unimodal and left-skewed, indicating a concentration of older individuals. The male age distribution is slightly bimodal and left-skewed, suggesting two modes within the distribution.

Differences between 2014 and 2015 Results:

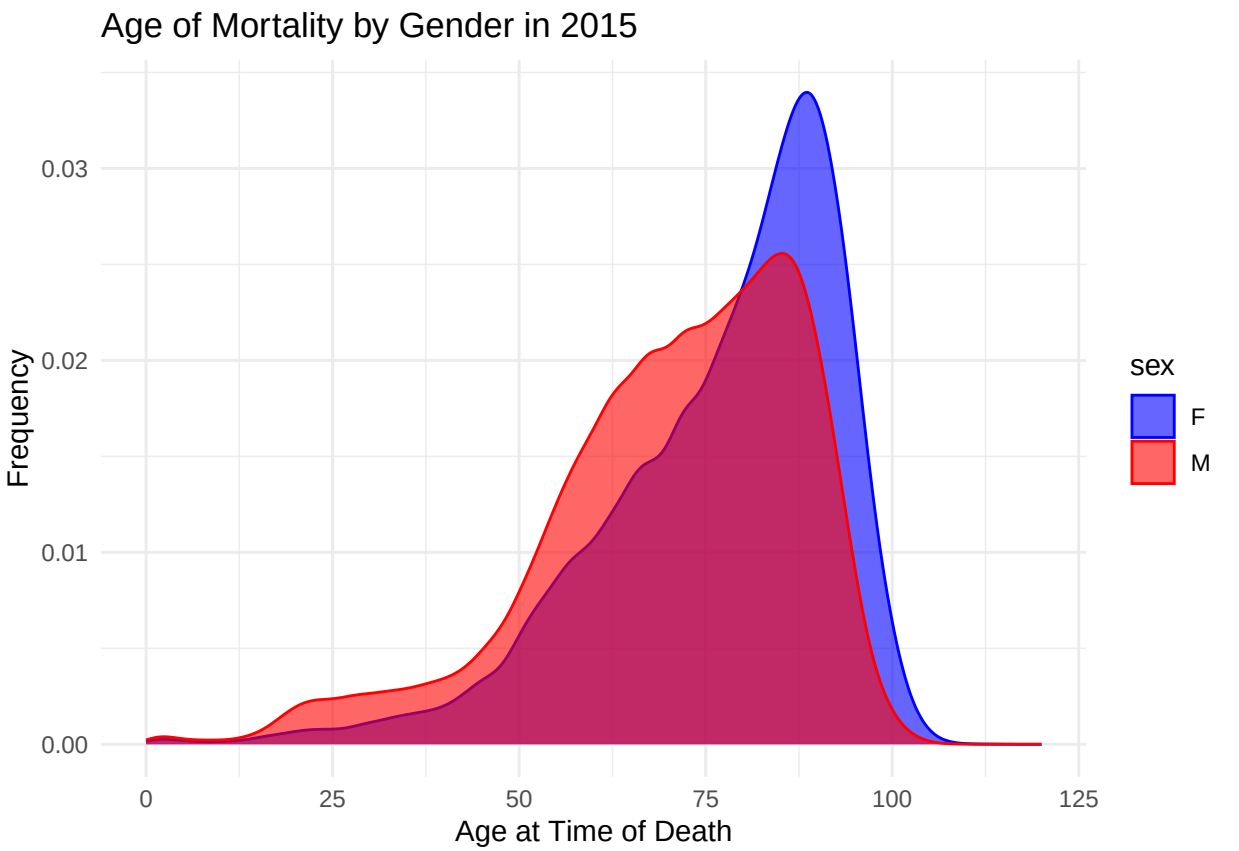
Comparing the age distributions of females and males in 2014 and 2015, several patterns remain consistent. Females continue to have higher mean and median ages compared to males, indicating longer life expectancy for women. Both data sets exhibit left-skewed distributions, suggesting a concentration of older individuals. The standard deviations and variances are similar between the two years, indicating consistent variability in age within each group. Overall, the 2015 data reaffirms the gender-based differences in life expectancy, with women living longer on average.

(b) What might explain the difference in averages?

Similar to 2014, there a few possible explanations for the differences in the averages in the 2015 data set:

- Biological differences can continue to play a role in the age disparities between genders, including genetic and hormonal factors that influence longevity.
- Differences in behaviors, healthcare access, and health outcomes may contribute to variations in life expectancy. For example, males are generally more prone to risky behaviors and certain health conditions.
- The differences in the mode and skewness may reflect the ongoing gender gap in mortality at different age groups, driven by factors like disease susceptibility and accidents.

```r
# Plot the distribution of age of mortality for both genders for a comprehensive view in 2015
sample_2015_cleaned %>%
  ggplot(aes(x = detail_age, color = sex, fill = sex)) +
  xlim(0, 120) +
  geom_density(alpha = 0.6) +
  scale_fill_manual(values = c("blue", "red")) +
  scale_color_manual(values = c("blue", "red")) +
  theme_minimal() +
  labs(title = "Age of Mortality by Gender in 2015", x = "Age", y = "Density") +
  ylab("Frequency") +
  xlab("Age at Time of Death")
```



In order to find the key differences between Females and Males, we found the summary statistics for each density plot.

```
# Find the summary statistics for the Density plot for Women/Men in 2015
sample_2015_cleaned_summarize_bysex = sample_2015_cleaned %>%
   group_by(sex) %>%
   summarize(mean_age_death=mean(detail_age), median_age_of_death=median(detail_age), sd_age_of_death=sd
             min_age_death=min(detail_age), max_age_of_death=max(detail_age), iqr_age_of_death=IQR(detai
```

Task 3: What are the most prevalent diagnoses for cause of death?

1) You would like to see which diseases, over the two years, are most prevalent. Consider the ICD10 code that describes the underlying cause of death (icd_code_10th_revision).

What are the top 5 most prevalent diseases?

The top five causes of death by order is as follows:

1. ICD-10-CM Diagnosis Code I25. 1: Atherosclerotic heart disease of native coronary artery.
2. ICD-10-CM Diagnosis Code: C34. 9 Malignant neoplasm: Bronchus or lung, unspecified.
3. ICD-10-CM Diagnosis Code: F03 Unspecified dementia, mild, without behavioral disturbance, psychotic disturbance, mood disturbance, and anxiety
4. ICD-10-CM Diagnosis Code: I21. 9 Acute myocardial infarction, unspecified.
5. ICD-10-CM Diagnosis Code: J44. 9 Chronic obstructive pulmonary disease, unspecified.

```r
# Count the top 5 most prevalent diagnoses and arrange in descending order
top_5_prevalent_diagnoses = sample_2014_2015_cleaned %>%
  count(icd_code_10th_revision) %>%
  arrange(desc(n)) %>%
  head(5)
```

2) You decide that the ICD 10 code is too granular and instead would like to look at a grouping of these codes. You opt for the *Elixhauser* comorbidity groupings from HCUP (https://www.hcup-us.ahrq.gov/toolssoftware/comorbidity/comorbidity.jsp). Read in the .csv called "Elixhauser.csv" that relates ICD codes to the Elixhauser comorbidity group and join this with the dataset used in 1 (above) to find the top 5 most common categories.

The top 5 most common comorbidity groups in the data are as follows:

1. Solid tumor without metastasis - 19,712 individuals in the data set
2. Chronic Pulmonary Disease - 10,821 individuals in the data set
3. Other Neurological Disorders - 10,660 individuals in the data set
4. Renal Failure - 2,977 individuals in the data set
5. Valvular Disease - 2,154 individuals in the data set

```r
# Read in the Elixhauser Data set
elixhauser_data = read_csv("Elixhauser.csv")
```

```
## Rows: 3493 Columns: 2
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (2): icd, disease
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

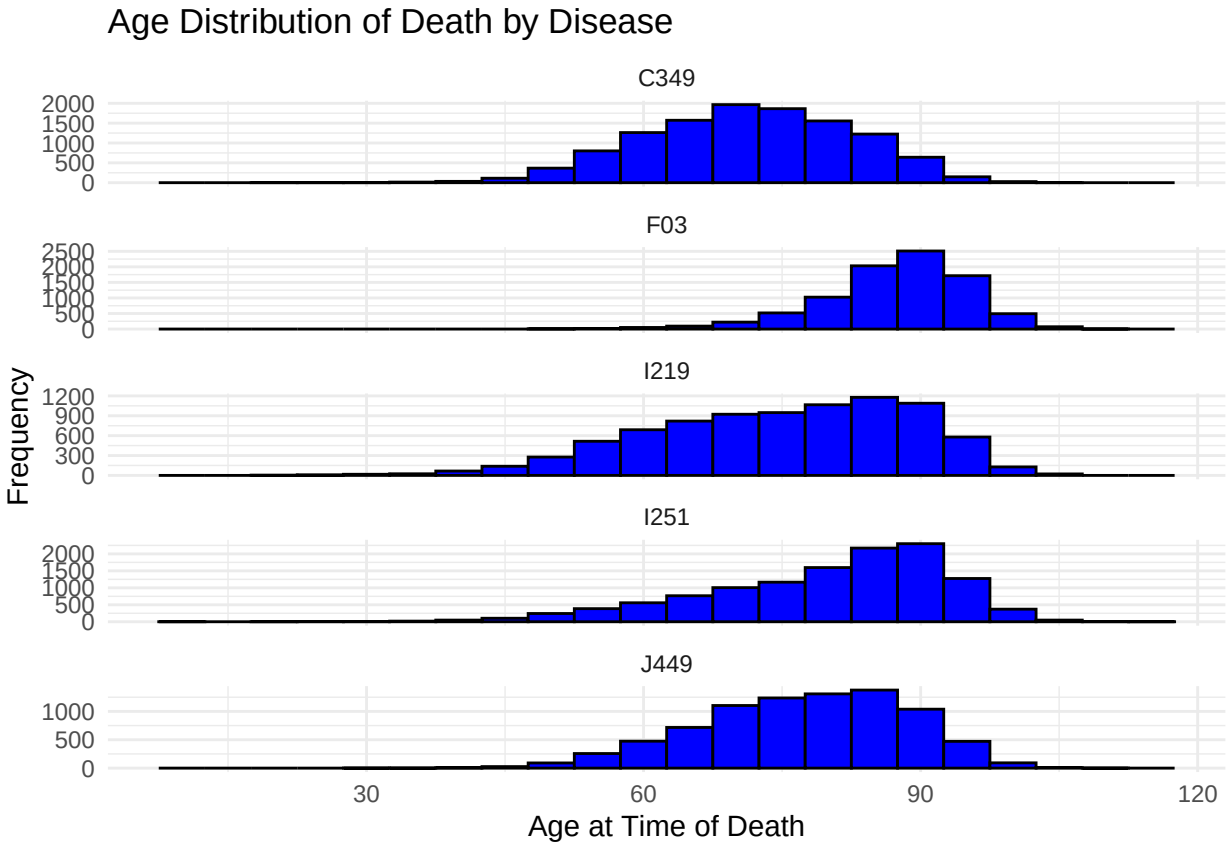
```r
# Combine the combined clean sample data set and elixhauser data using a left join
combined_eli_data = sample_2014_2015_cleaned %>%
  left_join(elixhauser_data, by = c("icd_code_10th_revision" = "icd"))

# Filter the top 5 most prevalent diagnoses and arrange in descending order
disease_top_5 = combined_eli_data %>%
  filter(!is.na(disease)) %>%
  count(disease) %>%
  arrange(desc(n)) %>%
  head(5)
```

3) Visualize the distribution of age of death for each of the top 5 disease categories found in 2 (above).

All the histograms exhibit a left-skewed distribution, which indicates that more individuals are dying at older ages due to these diseases, and fewer deaths are observed in younger age groups. This information provides a visual summary of the age distribution of death for these specific diseases, highlighting the increased prevalence of these diseases among the older population.

```r
## Plot the distribution of age of death for each of the top 5 comorbidity categories
top5_disease_categories = sample_2014_2015_cleaned %>%
  filter(icd_code_10th_revision %in% c("I251", "C349", "I219","J449", "F03"))
```

```r
top5_disease_categories %>%
  ggplot(aes(x = detail_age)) +
  geom_histogram(binwidth = 5, fill = "blue", color = "black") +
  facet_wrap(~icd_code_10th_revision, ncol = 1, scales = "free_y") +
  labs(title = "Age Distribution of Death by Disease",
       x = "Age at Time of Death", y = "Frequency") +
  theme_minimal()
```



Task 4: Create a recommendation for a target market/product

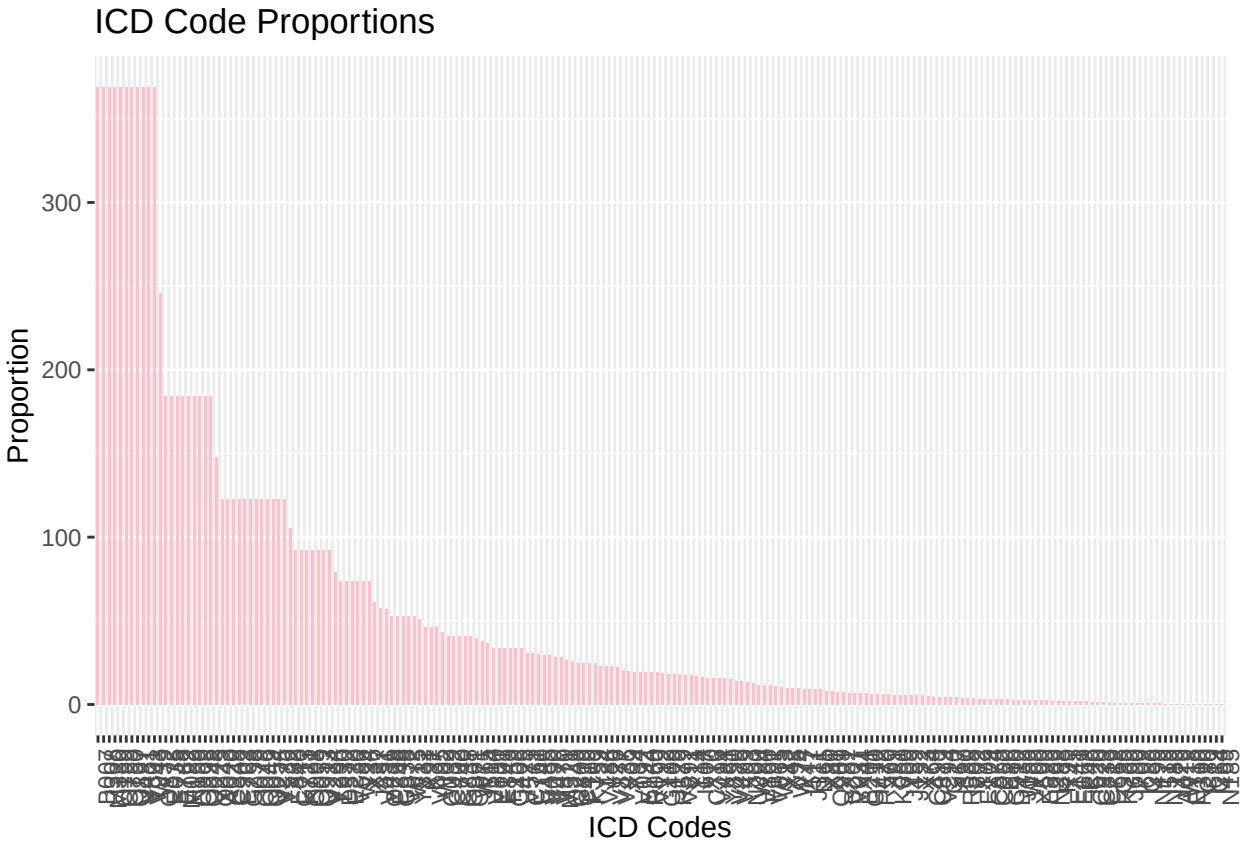
Building off your analyses in the previous tasks, consider a subset of the total population and a cause of death disproportionately impacting that population (e.g. young people dying from heart disease). Create a visualization contrasting this population with the population as a whole and explain how this informs your final target market/product recommendation.

In our analysis we looked to identify causes of mortality that disproportionately impact young females in the demographic 16-25 years of age. It is well know that females often experience disparities in care and we had interest in identifying a group that hopefully, through appropriate early interventions would have a large impact in years of life save, i.e. have a large impact on potential years of life added.

We first must calculate the prevalence rates of the disease related to mortality in this sample population. Next the prevalence of those same diseases in the total population are calculated. By dividing the sample prevalence by the population prevalence we are able demonstrate that there are a number of diseases that cause mortality in young females at a rate almost 400 times that of the general population. Those rates are graphically demonstrated in descending order.

By investigating the ICD codes associated with those disproportionate mortality rates, we found that death due to birth-related complications stood out in this population (i.e. ICD 10 codes starting O). While this is not surprising, it does suggest if a targeted campaign at early intervention for young mothers could make a large impact in mortality reduction. we should suggest early interventions with health screenings for both mother and the fetus as well as promotion of routine prenatal care could make a marked and lasting impact. Moreover, by saving young maternal lives there is not only a large number of life years added to the mother, but also likely enumerable impacts on the child.

```
# Find the number of young females, ages 18-25, in the combined data set
young_female_count <- sample_2014_2015_cleaned %>%
  filter(sex %in% c("F"), detail_age > 16, detail_age <25) %>%
  summarise(count = n())

# Find the number individuals of the full sample
count_size=sample_2014_2015_cleaned%>%
  summarise(count =n())

# Find the prevalence rate for young females (we multiplied this rate by 100,000 to get a rate per 100k
young_female_rate_per100K <- sample_2014_2015_cleaned %>%
  filter(sex %in% c("F"), detail_age > 16, detail_age <25) %>%
  group_by(icd_code_10th_revision) %>%
  summarise(prevalence_young_female = (n()/537)*100000) %>%
  arrange(desc(prevalence_young_female))

# Find the prevalence rate for the full sample population (we multiplied this rate by 100,000 to get a
sampled_rate_per100K <- sample_2014_2015_cleaned %>%
  group_by(icd_code_10th_revision) %>%
  summarise(prevalence_sample = n()/198275*100000) %>%
  arrange(desc(prevalence_sample))

# Join the 2 new data frames by the ICD codes
prevalence_data = sampled_rate_per100K %>%
  left_join(young_female_rate_per100K, by = c("icd_code_10th_revision" = "icd_code_10th_revision")) %>%
  filter(!is.na(prevalence_young_female))

# Find the proportion of diseases impacting young females vs the full sample population
prevalence_data_rate = prevalence_data %>%
  mutate(proportion = prevalence_young_female / prevalence_sample) %>%
  arrange(desc(proportion)) %>%
  head(200)

# Plot the prevalence of disease among young females ages 16 - 25
ggplot(prevalence_data_rate, aes(x = reorder(icd_code_10th_revision, -proportion), y = proportion)) +
  geom_bar(stat = "identity", fill = "pink", width = 0.5) +
  labs(title = "ICD Code Proportions", x = "ICD Codes", y = "Proportion") +
  theme(axis.text.x = element_text(angle = 90, hjust = 1))
```


ICD Code Proportions
Proportion
300
200
100
0
ICD Codes