

# Project

2023-03-08

## *Testing the Mysteries of Food Effects*

Report by Abena Boateng and Sedem Kakrada

### Introduction

The dataset 'cancer' is provided by the CDC's division "United States Cancer Statistics" at <https://www.cdc.gov/cancer/uscs/> (<https://www.cdc.gov/cancer/uscs/>). This dataset includes information by State and Year regarding the Death rate of cancer and the Incidence Rate both on a scale of 100,000 people.

There are 400 observations with 4 variables. Each row corresponds to a State in a specified Year:

```
#rows of cancer  
nrow(cancer)
```

```
## [1] 400
```

```
cancer %>%  
  summarize(n_StatebyYear = n())
```

```
## # A tibble: 1 × 1  
##   n_StatebyYear  
##           <int>  
## 1             400
```

The data set 'gmo' was obtained the USDA at <https://www.ers.usda.gov/data-products/adoption-of-genetically-engineered-crops-in-the-u-s/> (<https://www.ers.usda.gov/data-products/adoption-of-genetically-engineered-crops-in-the-u-s/>). The dataset set provides information on the GMO adoption by the Year and state also providing the type of engineering and crop.

There are 2231 observations with 5 variables. Each row corresponds to a State in a specified Year:

```
#rows of gmo  
nrow(gmo)
```

```
## [1] 2231
```

```
gmo %>%  
  summarize(n_StatebyYear = n())
```

```
## # A tibble: 1 × 1
##   n_StatebyYear
##           <int>
## 1           2231
```

The data set 'policy' is a data set of three different types of food policy bills from “Comparative Agendas” at <https://www.comparativeagendas.net/tool> (<https://www.comparativeagendas.net/tool>). The three different types of bills are Food Inspection & Safety, Marketing & Promotion Agriculture, and Animal and Crop Disease.

There are 17 observations with 4 variables. Each row corresponds to a specified Year:

```
# rows of policy
nrow(policy)
```

```
## [1] 17
```

```
policy %>%
  summarize(n_Year = n())
```

```
## # A tibble: 1 × 1
##   n_Year
##   <int>
## 1     17
```

```
##Renaming
policy <- policy %>%
  rename(safetypolicy = 'US: Bills #Agriculture: Food Inspection & Safety')
policy <- policy %>%
  rename(marketingpolicy = 'US: Bills #Agriculture: Marketing & Promotion')
policy <- policy %>%
  rename(diseasepolicy = 'US: Bills #Agriculture: Animal and Crop Disease')
```

## Research Questions

**Do genetically engineered crops seem to have an influence on the cancer incidence rate and death rate in the United States?**

**What is the relationship between the number of food policy bills and the and cancer incidence and death rates by year?**

Based on the speculation of GMOs to health effects, we wanted to explore its relationship to one type of chronic condition, cancer, that is often linked to nutrition and lifestyle factors. In also considering the effect of food and nutrition on the development of chronic conditions, we are also going to look at the regulations through legislation on the effect of food production quality which could impact cancer incidence. We expect to find that the increase of GMO could correlate to an increase in cancer incidence and death rates as well as policy rates increasing being effective at protecting the quality of nutrients in the food market that could lower cancer rates.

# Tidying

```
#pivoting the values of gmo type
gmo <- pivot_wider(gmo,
  names_from = Type,
  values_from = Value)
```

To make the final data set more tidy, a `pivot_wider` function is used on the `Type` column in order to have each engineering type with its own column and its corresponding values of adoption relating the State and Year.

## Joining/Merging

```
#joining of three datasets
project <- inner_join(gmo, policy, by = c('Year'))
project <- inner_join(project, cancer, by = c('Year', 'State'))
```

```
#observations of each dataset
nrow(cancer)
```

```
## [1] 400
```

```
nrow(gmo)
```

```
## [1] 989
```

```
nrow(policy)
```

```
## [1] 17
```

```
#rows dropped/added
nrow(project) - nrow(cancer)
```

```
## [1] -256
```

```
nrow(project) - nrow(gmo)
```

```
## [1] -845
```

```
nrow(project) - nrow(policy)
```

```
## [1] 127
```

There were 400 observations in cancer dataset, 989 observations in the gmo dataset, and 17 observations in the policy dataset. The IDs in common between the gmo and cancer datasets were Year and State while the policy dataset only had the ID year as a common variable to the two other datasets. There are 2 unique IDs in the cancer dataset, Incidence Rate and the Death Rate. In the gmo dataset there are 3 unique IDs, Crop, herbicide, insectresistant, and stacked gene. Lastly, in the policy dataset the 3 unique IDs include safetypolicy, marketingpolicy, and diseasepolicy. After joining the datasets there were no IDs that are left out of the final dataset. When the datasets were merged the cancer dataset lost 257 rows, the gmo dataset lost 845 rows and the policy dataset gained 127 rows. —————

# Wrangling

## Dplyr Functions

```
#finding the top 3 death rate with corresponding State
project %>%
  select(State, Deaths) %>%
  top_n(3, Deaths)
```

```
## # A tibble: 3 × 2
##   State      Deaths
##   <chr>      <dbl>
## 1 California  58412
## 2 California  59629
## 3 California  59515
```

```
project %>%
  select(State, Deaths) %>%
  top_n(-3, Deaths)
```

```
## # A tibble: 4 × 2
##   State      Deaths
##   <chr>      <dbl>
## 1 North Dakota  1302
## 2 North Dakota  1253
## 3 North Dakota  1302
## 4 North Dakota  1253
```

The three highest death rates correspond to the state of California. As a death rate out of 100,000 people this likely to be influenced by population density and because California is significantly more populated than the state with the lowest death rate with more urban areas and modern lifestyles linked to the increase of chronic conditions.

```
#combining the policy amount by year for new variable
project2 <- project %>%
  group_by(Year, State, Crop, Rate, Deaths) %>%
  mutate(totalpolicy = sum(safetypolicy, marketingpolicy, diseasepolicy)) %>%
  mutate(totalgmo = sum (herbicide, insectresistant, stackedgene, na.rm = TRUE))
```

Combining the different types of policies into a summation of policies proposed each year into a new variable, totalpolicy. The same is done with the different types of genetic engineering adoption values into the variable totalgmo. The creation of these variables make it easier for establishing relationships between the variables.

```
#frequency of gmo types
project %>%
  filter(State == 'Texas') %>%
  group_by(Year) %>%
  arrange(desc(Rate))
```

```
## # A tibble: 8 × 11
## # Groups:   Year [4]
##   Year State Crop herbi...1 insec...2 stack...3 safet...4 marke...5 disea...6 Rate Deaths
##   <dbl> <chr> <chr>    <dbl>    <dbl>    <dbl>    <dbl>    <dbl>    <dbl> <dbl>
## 1  2005 Texas corn      42      21      9      26      9      18  179.  34291
## 2  2005 Texas cott...  35      14     14     26      9      18  179.  34291
## 3  2014 Texas corn      17      12     62      7      4       5  153.  38847
## 4  2014 Texas cott...  15       4     74      7      4       5  153.  38847
## 5  2015 Texas corn      12      10     67     25      1       7  149.  39121
## 6  2015 Texas cott...  11       5     75     25      1       7  149.  39121
## 7  2016 Texas corn      11       8     71      3      1       0  148.  40195
## 8  2016 Texas cott...  11       4     75      3      1       0  148.  40195
## # ... with abbreviated variable names 1herbicide, 2insectresistant, 3stackedgene,
## # 4safetypolicy, 5marketingpolicy, 6diseasepolicy
```

Taking a closer look at a specific state, Texas, we can see that the year 2005 has the highest cancer incidence rates. Although this year had the highest incidence rate the highest death rate is in the most recent year 2016 suggesting there is some other effector that has been introduced over the years that could contribute to the steady increase death rates

```
#mean cancer rate by state and year
project2 %>%
  group_by(State) %>%
  summarize(mean = mean(Rate)) %>%
  arrange(desc(mean))
```

```
## # A tibble: 20 × 2
##   State      mean
##   <chr>    <dbl>
## 1 Mississippi 195.
## 2 Arkansas    189.
## 3 Tennessee   189.
## 4 Alabama     183.
## 5 Indiana     182.
## 6 Ohio        181.
## 7 Missouri    179.
## 8 Michigan    175.
## 9 Illinois    173.
## 10 North Carolina 172.
## 11 Georgia     170.
## 12 Kansas      169.
## 13 Iowa        168.
## 14 Wisconsin   165.
## 15 South Dakota 164.
## 16 Nebraska    162.
## 17 Texas       157.
## 18 Minnesota   156.
## 19 North Dakota 155.
## 20 California  149.
```

```
project2 %>%
  select(totalgmo, State) %>%
  filter(State == 'Tennessee')
```

```
## Adding missing grouping variables: `Year`, `Crop`, `Rate`, `Deaths`
```

```
## # A tibble: 4 × 6
## # Groups:   Year, State, Crop, Rate, Deaths [4]
##   Year Crop    Rate Deaths totalgmo State
##   <dbl> <chr>  <dbl> <dbl>    <dbl> <chr>
## 1  2005 cotton  210.  12995      96 Tennessee
## 2  2014 cotton  184.  14172      99 Tennessee
## 3  2015 cotton  180.  14214      99 Tennessee
## 4  2016 cotton  180.  14450      98 Tennessee
```

After grouping by the state, we can see that the state Mississippi had the highest average of cancer incidents over the four years being studied with a mean of 194.525. California had the lowest average with a value of 148.875. When looking at the total GMO adoption in the leading incidence rate state Tennessee, we can see that in the year with the highest rate the amount of GMOs adopted was less than the year with the lowest incidence rate.

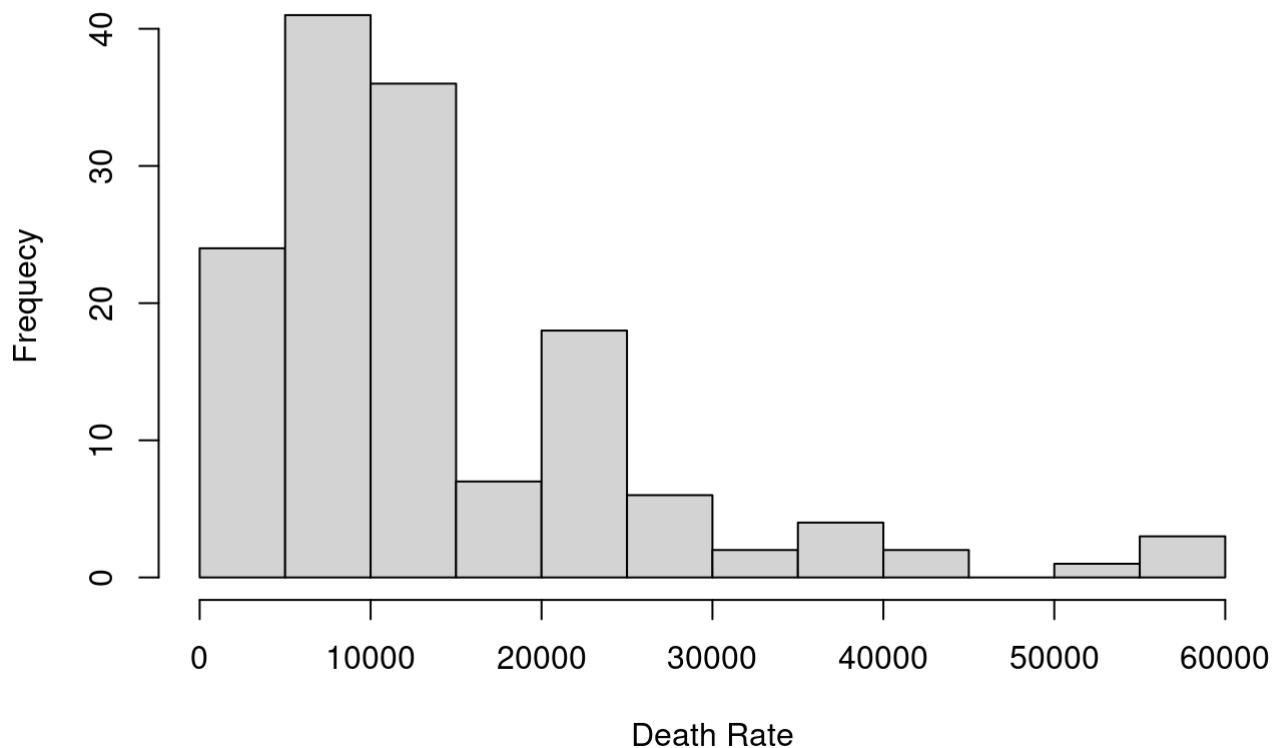
# Summary Statistics

```
#summary stats of Deaths
summary(project$Deaths)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      1253   6414   11408   13997   19666   59629
```

```
hist(project$Deaths, xlab = 'Death Rate', ylab = 'Frequency', main = 'Distribution of Deaths by Cancer')
```

## Distribution of Deaths by Cancer



```
IQR(project$Deaths)
```

```
## [1] 13251.5
```

Due to the skewed distribution of Deaths, the median deaths per 100,000 incidents to report is 11408 with the IQR reporting at 13251.5. While the highest Death Rate can be nearly 60,000 people, the lowest death rate reported in this data set and utilized for determining any correlations to GMO adoption and food policy.

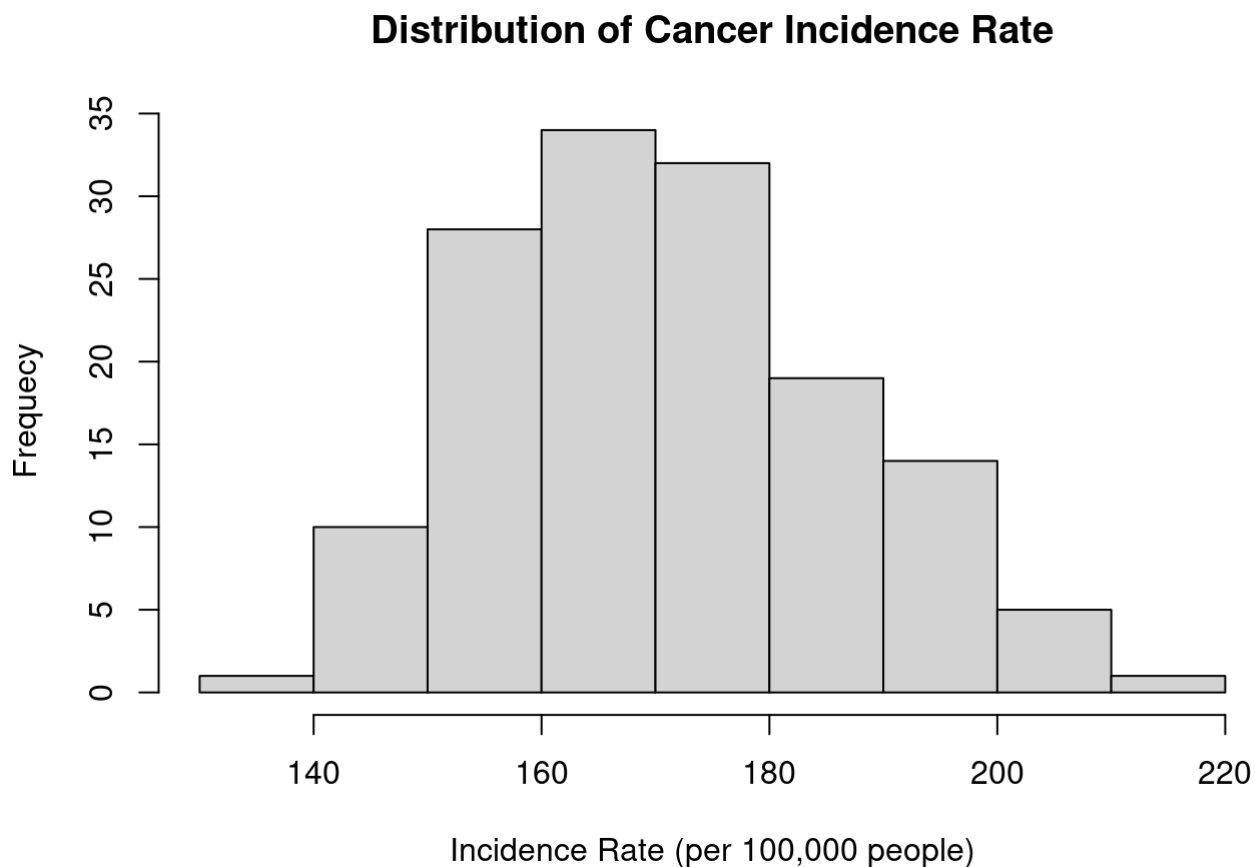
```
#summary stats of Cancer incidence rate by state and year
summary(project$Rate)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    139.7   159.6   169.2   171.7   182.7   210.3
```

```
sd(project$Rate)
```

```
## [1] 15.94505
```

```
hist(project$Rate, xlab = 'Incidence Rate (per 100,000 people)', ylab = 'Frequency', main = 'Distribution of Cancer Incidence Rate')
```



The cancer incidence rate used for this analysis report has a normal distribution with an average of 171.7 incidents of cancer per 100,000 people. This rate utilized that data from the select states focused on throughout the report rather than a national average and the mean within each state likely have a varied value of about 16.06 standard deviations.

```
#summary stats of Food related policies by year
summary(project2$totalpolicy)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##     4.0    13.0    24.5    26.5    38.0    53.0
```

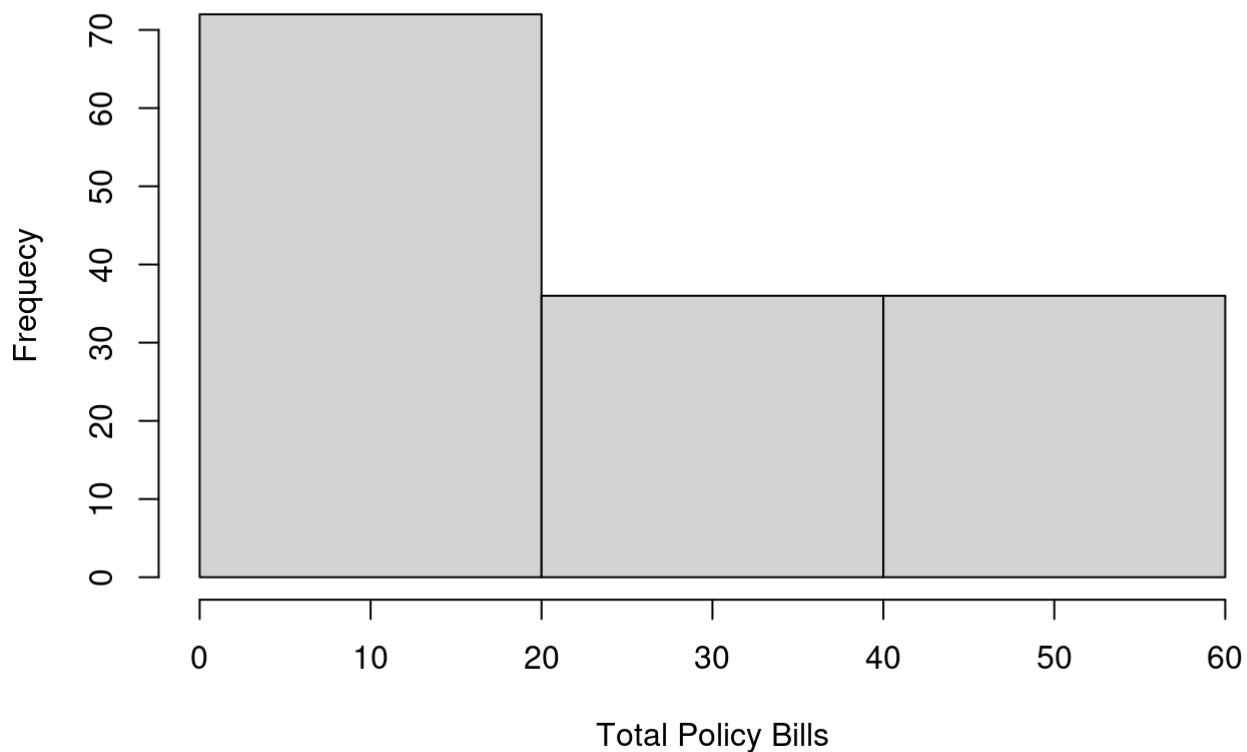


```
IQR(project2$totalpolicy)
```

```
## [1] 25
```

```
hist(project2$totalpolicy, xlab = 'Total Policy Bills', ylab = 'Frequency', main = 'Distribution of Total Policy Bills', breaks = 3)
```

## Distribution of Total Policy Bills



The distribution of each individual policy type resulted in a spaced histogram due to a noncontinuous scale of values, in order to produce a better visualization the combination of all policy types is produced. Once the breaks are adjusted the distribution of food policies over the selected year produce a right skewed graph with a median of 24.5 policies presented per year. The total policy bills going through the legislative branch each year has an interquartile range on 25 bills.

```
#Count of each State  
table(project$State)
```

```
##
```

|    | Alabama     | Arkansas     | California | Georgia        | Illinois     |
|----|-------------|--------------|------------|----------------|--------------|
| ## | 4           | 8            | 4          | 4              | 8            |
|    | Indiana     | Iowa         | Kansas     | Michigan       | Minnesota    |
| ## | 8           | 8            | 8          | 8              | 8            |
|    | Mississippi | Missouri     | Nebraska   | North Carolina | North Dakota |
| ## | 8           | 12           | 8          | 4              | 8            |
|    | Ohio        | South Dakota | Tennessee  | Texas          | Wisconsin    |
| ## | 8           | 8            | 4          | 8              | 8            |

This table of frequency relates the amount of data this report uses for each state. Missouri is the most frequent state with a frequency of 12, meaning that for the data joined and utilized for the exploration of these topics this state had the most rows of data. The least frequent states with the least amount of data are Alabama, California, Georgia, North Carolina, and Tennessee each with a frequency of 4 observations. This difference in data collection is amount is likely due to the lack of federal data collected for GMO seed use by the USDA. The estimates given about the GMO use values is independently collected by the International Service for the Acquisition of Agri-biotech Applications in which some states may not have data for.

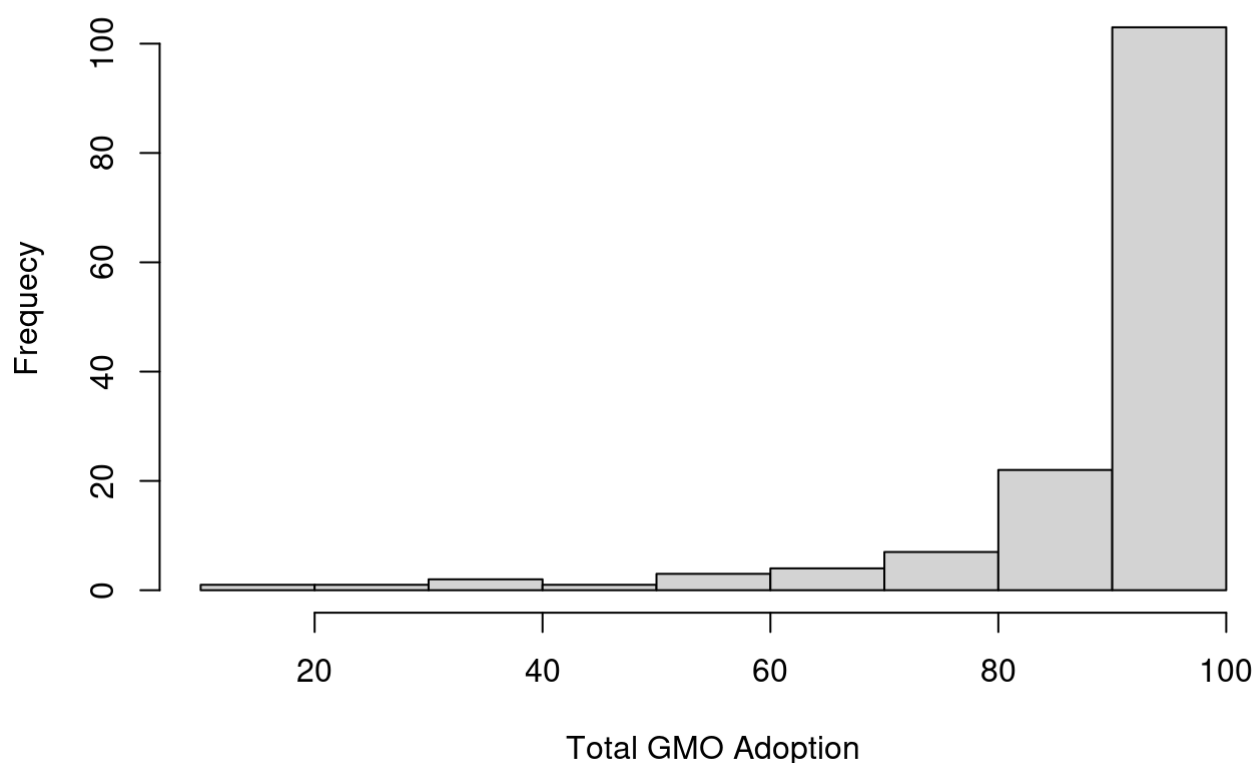
```
#count of each gmo crop type
table(project$Crop)
```

```
##
##   corn  cotton soybean
##    52    36    56
```

The data set of the GMO crops focuses on three different crop sources in which soybean is the most frequent with a count of 56, followed by corn with a count of 52, and cotton with a count of 36. These frequencies in the data set do not directly align with the national trend of corn being the most popularly grown crop.

```
#summary stats of totalgmo
hist(project2$totalgmo, xlab = 'Total GMO Adoption', ylab = 'Frequency', main = 'Distribution of Total GMO Adoption' )
```

## Distribution of Total GMO Adoption



```
summary(project2$totalgmo, na.rm = TRUE)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      18.00   89.75   93.00   89.15   96.00   99.00
```

```
IQR(project2$totalgmo, na.rm = TRUE)
```

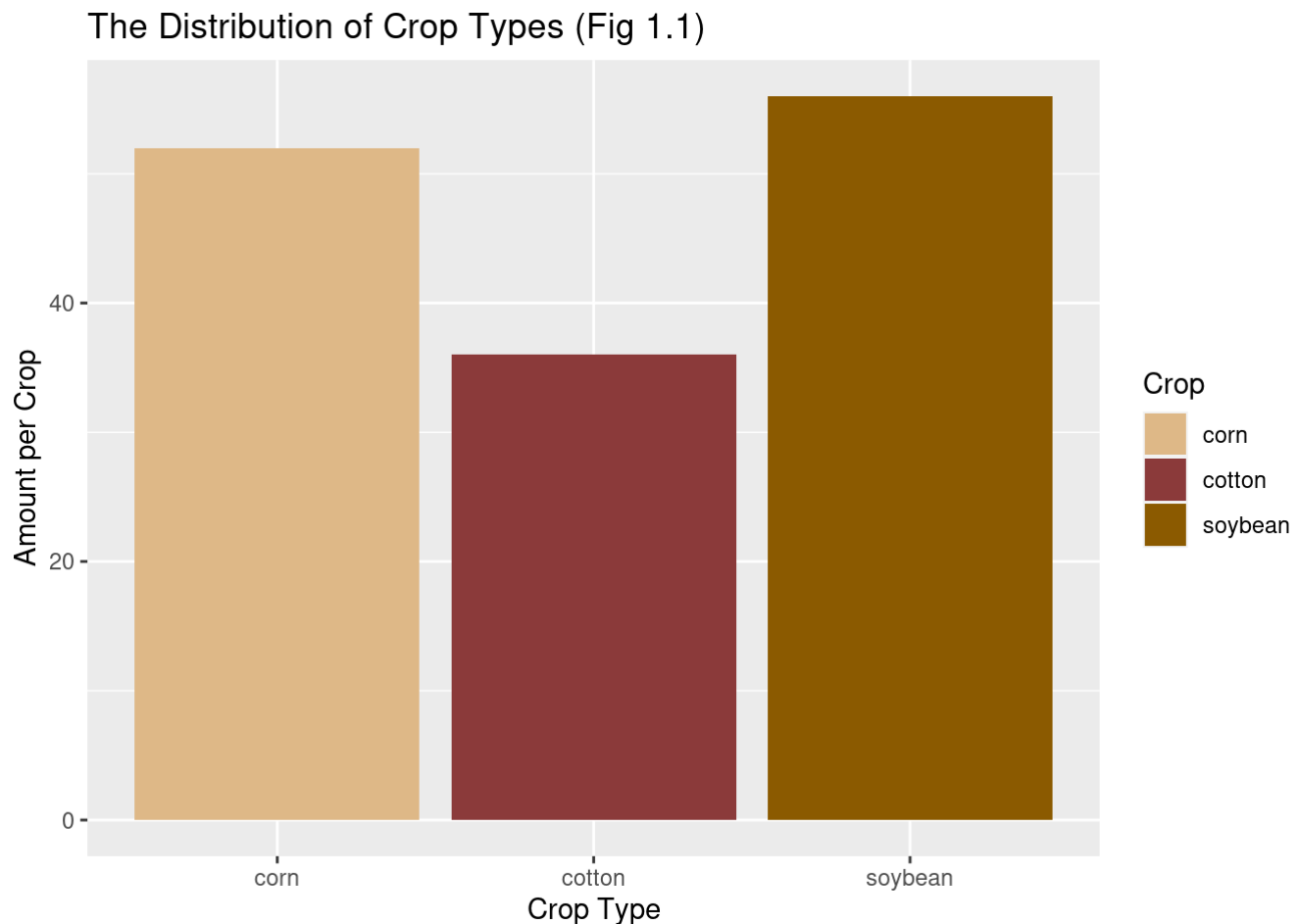
```
## [1] 6.25
```

The distribution of the total amount of GMO Adoptions is highly left skewed with a median of 93.00 adoptions and an IQR of 6.25. The skew of the graph is due to the widening acceptance of GMO adoption for mass food production in which more state are utilizing it.

## Visualizing

### Univariate Graphs

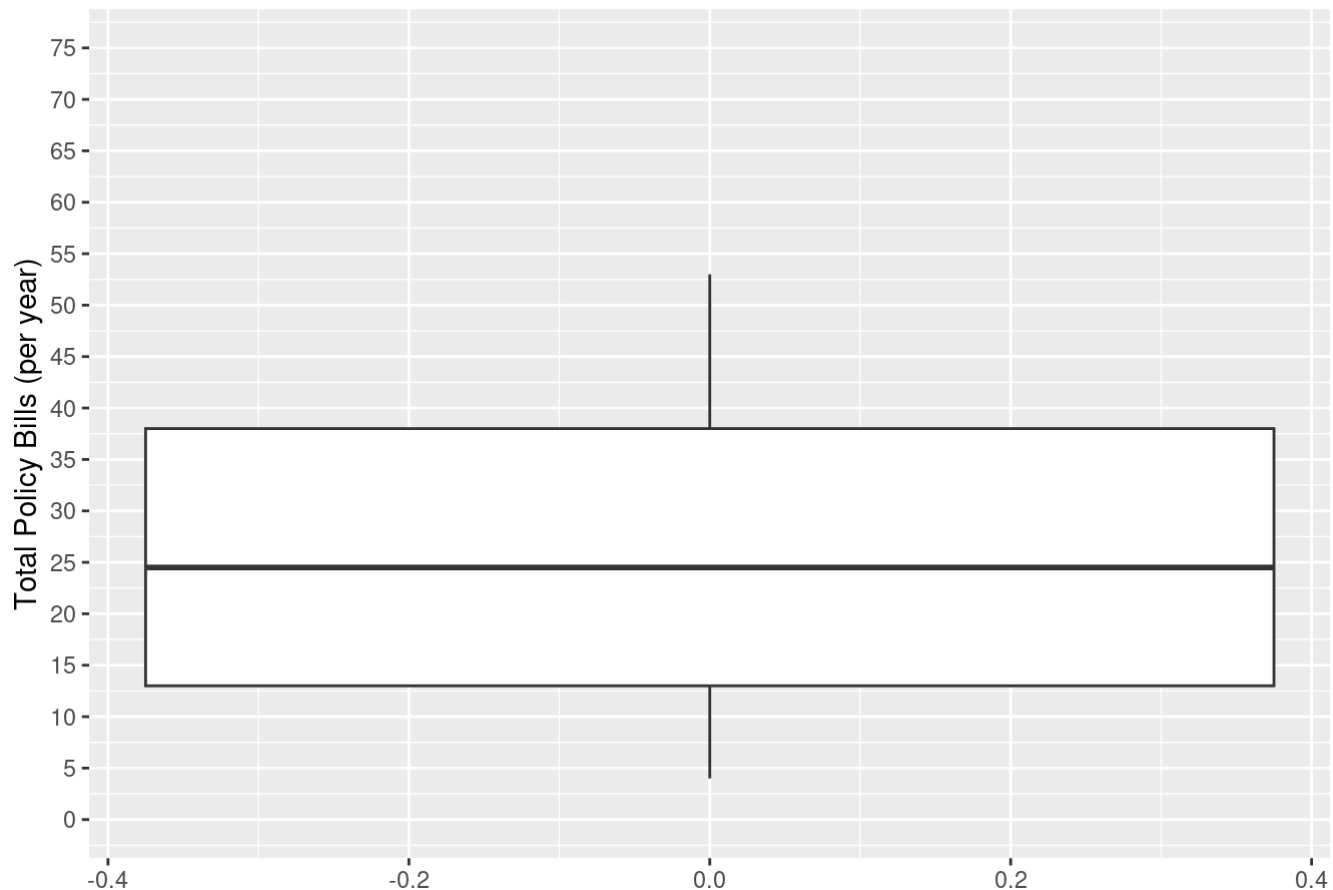
```
# bar graph of crop types
project %>%
  ggplot(aes(x = Crop, fill = Crop)) +
  geom_bar() +
  scale_fill_manual(values=c("burlywood", "indianred4", "orange4")) +
  labs(title= "The Distribution of Crop Types (Fig 1.1)", x = 'Crop Type', y = 'Amount p
er Crop')
```



A uni variate bar plot was created to visualize the distribution between the different GMO crop types. After a close observation, it is evident that the genetic engineering of the soybean crop type has the highest value yielding around 58 crops while cotton crop type has the lowest value yielding around 36 crops.

```
# box plot distribution of total policy bills within a year
project2 %>%
  ggplot(aes(y = totalpolicy)) +
  geom_boxplot() +
  labs(title = 'Distribution of the Total Policy Bills (Fig 1.2)', y = 'Total Policy Bi
lls (per year)') +
  scale_y_continuous(breaks=seq(0,75,5), limits=c(0, 75))
```

Distribution of the Total Policy Bills (Fig 1.2)

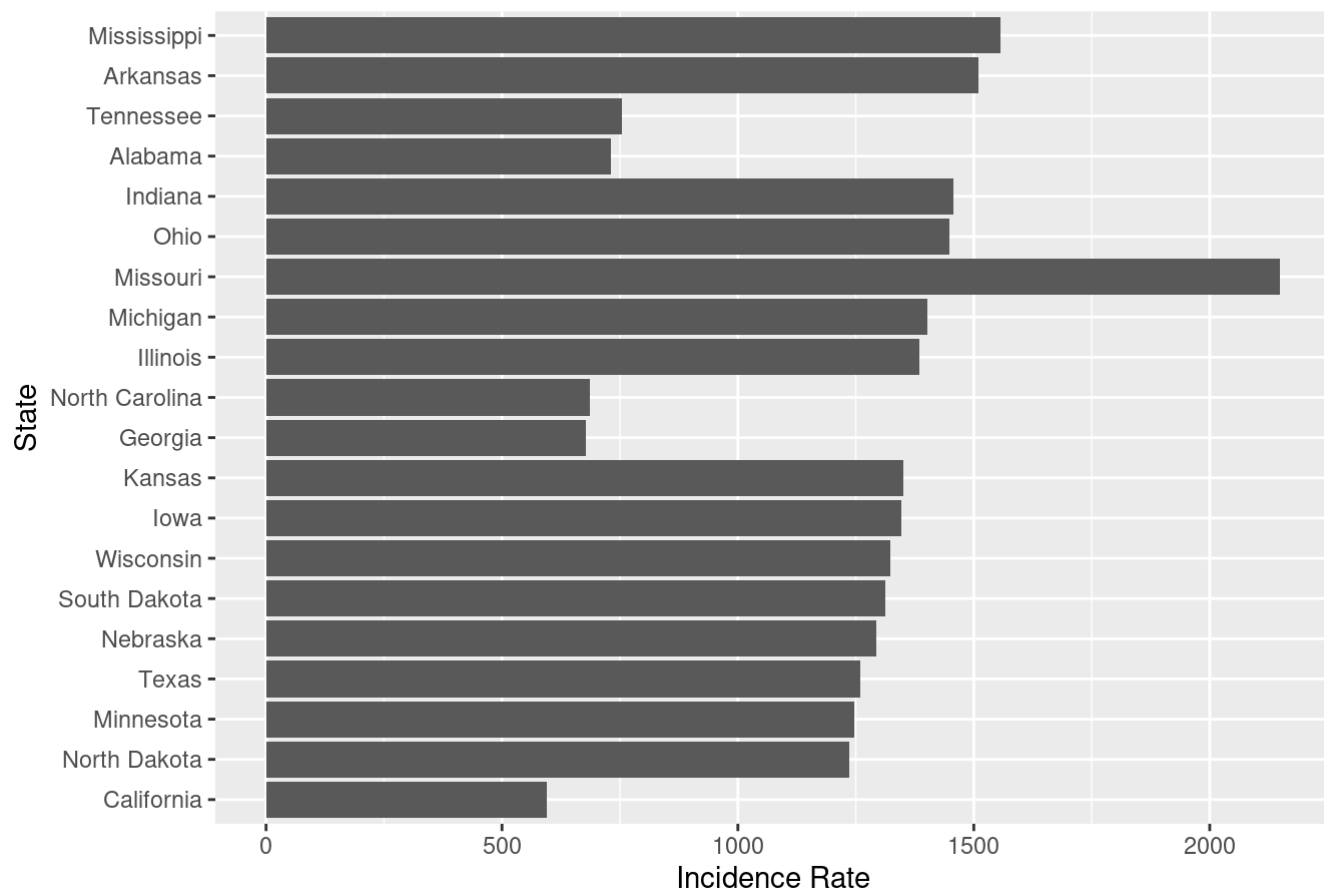


By means of a boxplot picture above, an analysis of the distribution of total policy bills passed appears to be symmetrical. To further describe the number of bills passed per year we could use mean and standard deviation as the boxplot does not appear to be skewed or contain many outliers.

### Bivariate Graphs

```
# bar graph of state with cancer rate
project2 %>%
  ggplot(aes(x = Rate, y = reorder(State, Rate))) +
  geom_bar(stat = 'identity') +
  labs(title = 'Cancer Incidence Rate based on each State (per 100,000 people) (Fig 1.
3)', x = 'Incidence Rate', y = 'State') +
  theme(legend.position = "none") +
  scale_x_continuous(breaks=seq(0,3000,500))
```

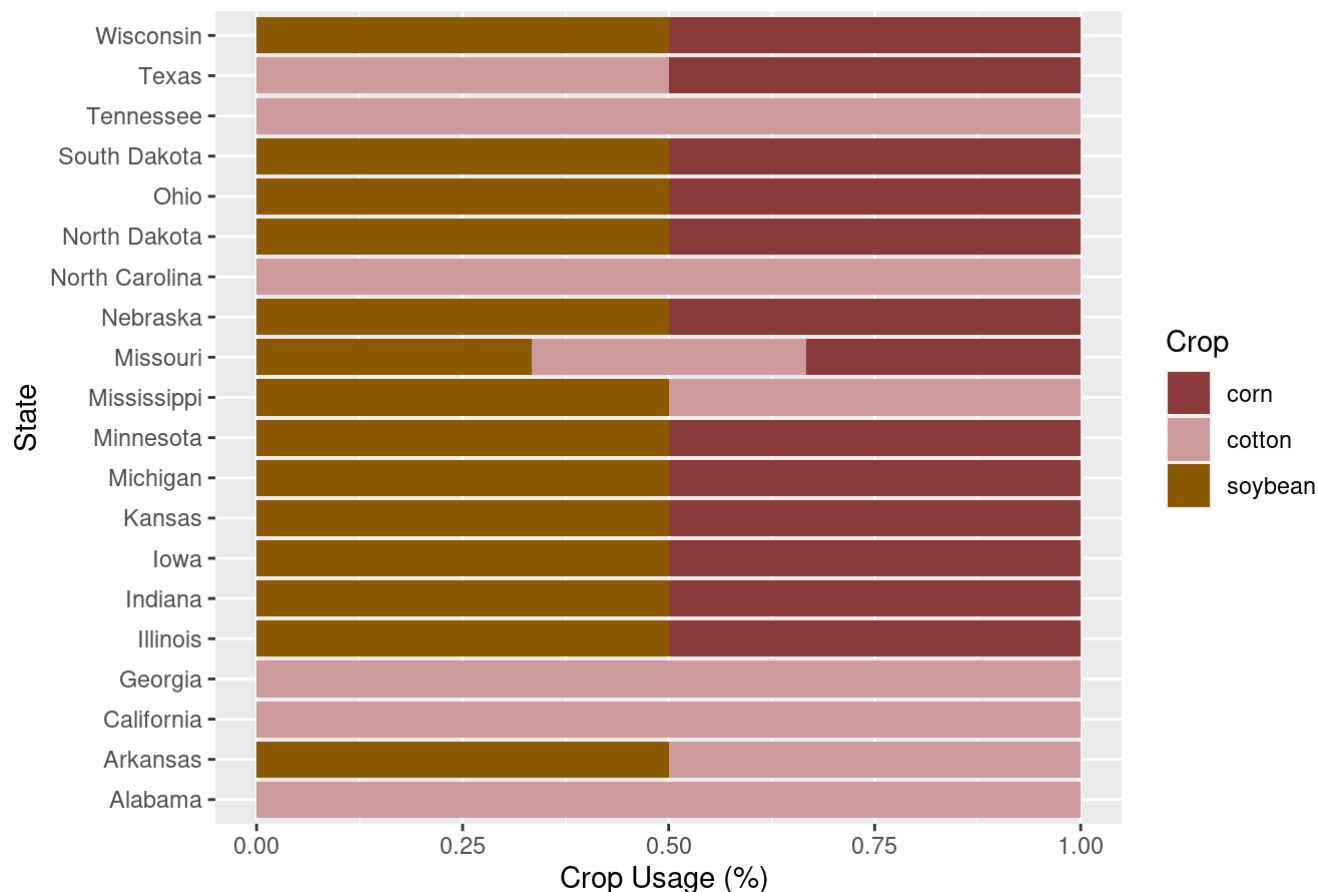
Cancer Incidence Rate based on each State (per 100,000 people) (Fig



In the figure above there a barplot is created to visualize the categorical variable of State and the numeric variable of Cancer Incidence Rate. As pictured above there are lots of states with very low incidence rates such as California, Alabama, and Georgia. However on the other case there are certain states displaying high amounts of Incidence Rates such as Missouri, Mississippi, and Arkansas.

```
# type of gmo crop usage by state
project2 %>%
  ggplot(aes(y = State, fill = Crop)) +
  geom_bar(position = 'fill') +
  labs(title = "GMO Crop Usage for each State (Fig 1.4)", y = 'State', x = 'Crop Usage (%)') +
  scale_fill_manual(values=c("indianred4", "rosybrown3", "orange4"))
```

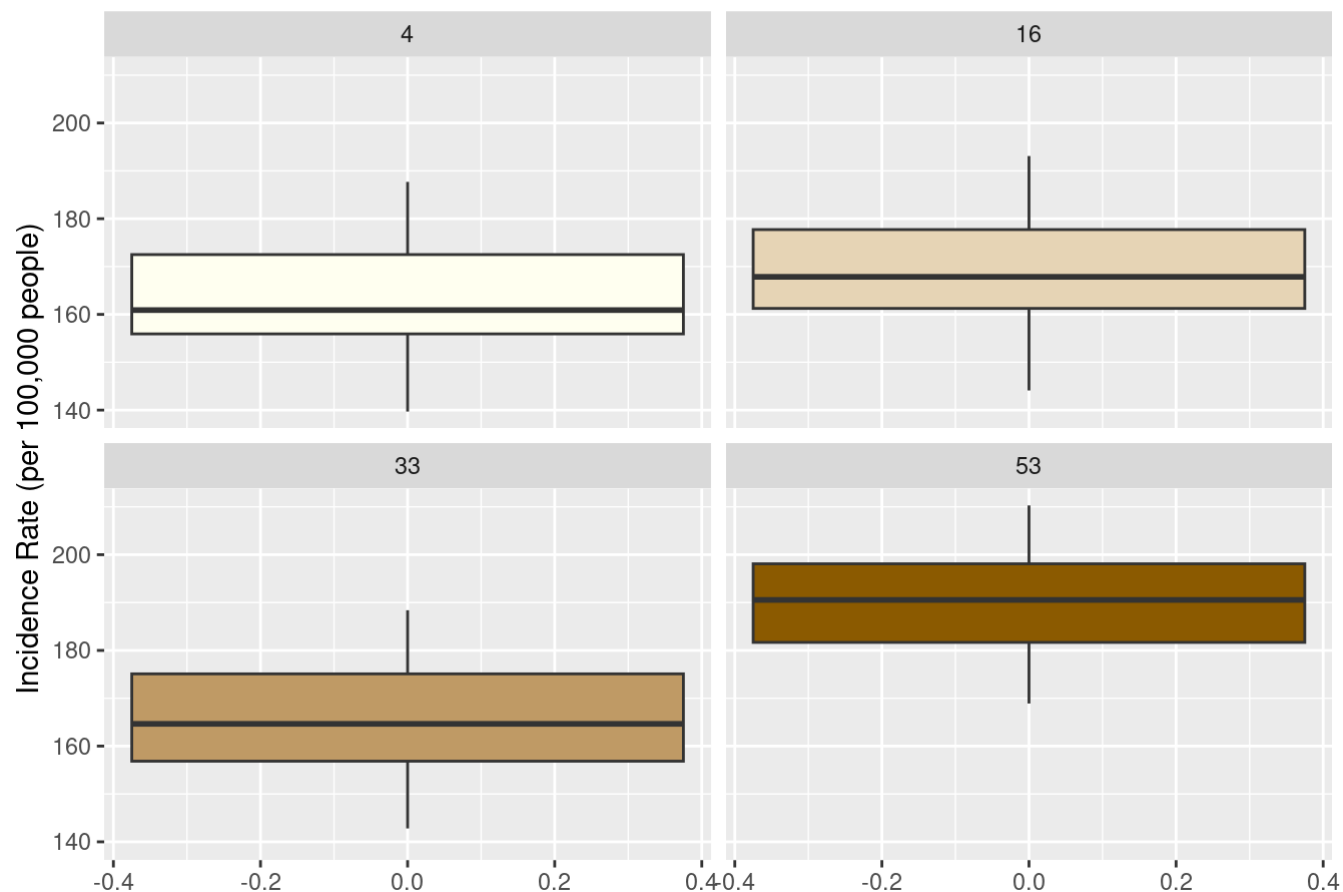
GMO Crop Usage for each State (Fig 1.4)



As pictured above a grouped barplot aided in the visualization of Crop Usage Percentage for each State. After observing the barplot it is evident that most states in our dataset rely on solely two types GMO crops being combinations of soybean and cotton, cotton and corn, or soybean and corn. With this in mind, the visualization show that for these combinations *there is a equal amount of GMO crop usage for each type*. In spite of this, overall it is evident that if a GMO crop type were to be solely used it would be cotton.

```
# cancer rate distribution by crop type
project2 %>%
  ggplot(aes(y = Rate, fill = totalpolicy)) +
  labs(title = 'Distribution of Cancer Incidence Rates by Policy Bills (Fig 1.5)', y =
'Incidence Rate (per 100,000 people)') +
  geom_boxplot() +
  facet_wrap(~totalpolicy) +
  scale_fill_continuous(high = 'orange4', low = 'ivory') +
  theme(legend.position = 'none')
```

Distribution of Cancer Incidence Rates by Policy Bills (Fig 1.5)



In the series of box plots pictured above the relationship between the amount of policy bill proposed in the corresponding years and Cancer incidence rates is faceted to showcase the similarities and differences. The earliest year in this study is 2005 which corresponds to the highest policy proposals, 53 bills, and has the highest valued distribution of cancer incidence rate. On the contrary, the latest year studied in this report, 2016, has a total policy bill value of 4 bills and has the lowest cancer incidence rate distribution.

### Trivariate Graphs

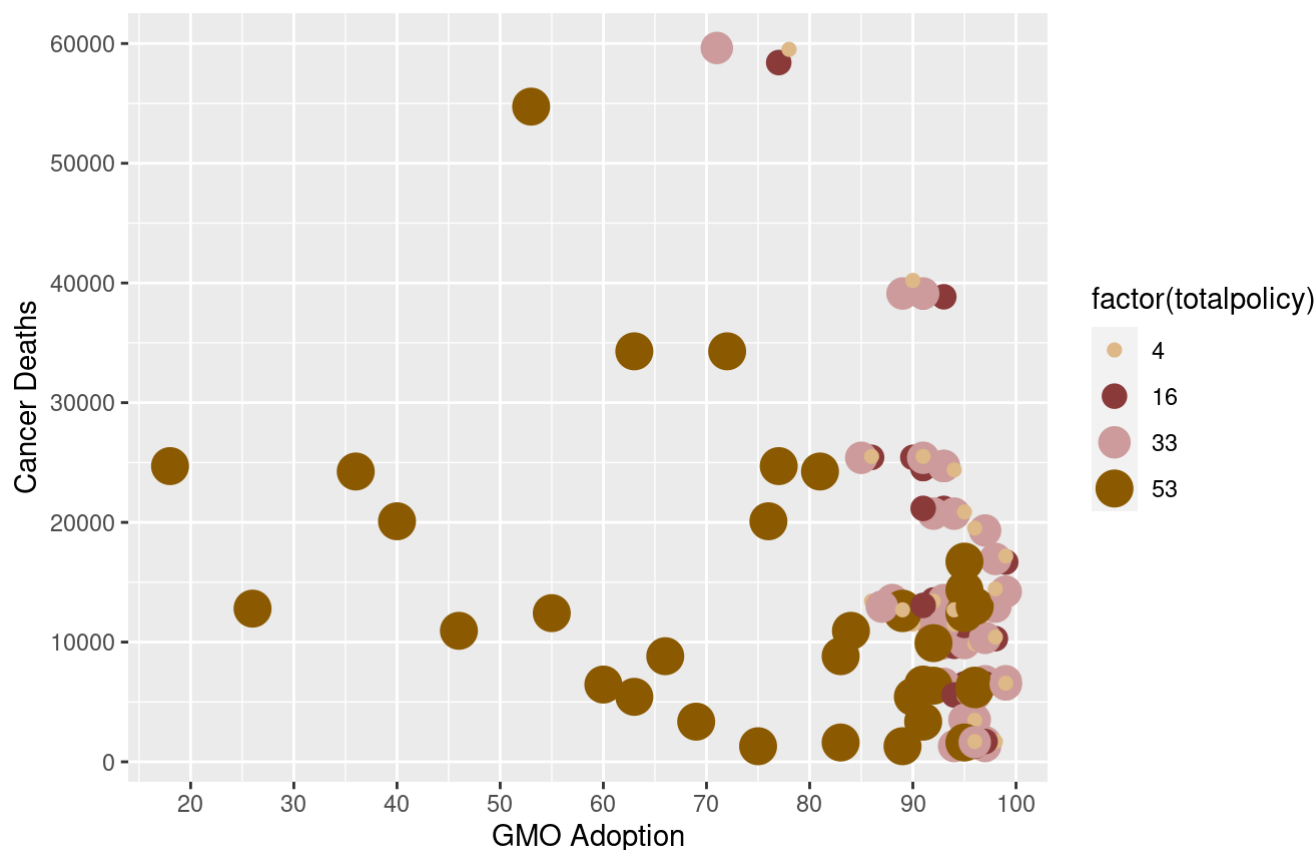
```
# correlation between gmo adoption and cancer death rate and total policy number
project2 %>%
  ggplot(aes(x = totalgmo, y = Deaths, size = factor(totalpolicy))) +
  labs(title = 'Relationship between GMO Adoption and Cancer Death Rate (Fig 1.6)', subtitle = 'Cancer Death Rate vs. GMO Adoption vs.Total Policy Number', x = 'GMO Adoption', y = 'Cancer Deaths' )+
  geom_point(aes(colour = factor(totalpolicy)), ) +
  scale_y_continuous(breaks=seq(0,70000,10000))+
  scale_x_continuous(breaks=seq(0,150,10)) +
  scale_color_manual(values=c("burlywood", "indianred4", "rosybrown3", "orange4"))
```

```
## Warning: Using size for a discrete variable is not advised.
```



## Relationship between GMO Adoption and Cancer Death Rate (Fig 1.6)

Cancer Death Rate vs. GMO Adoption vs. Total Policy Number



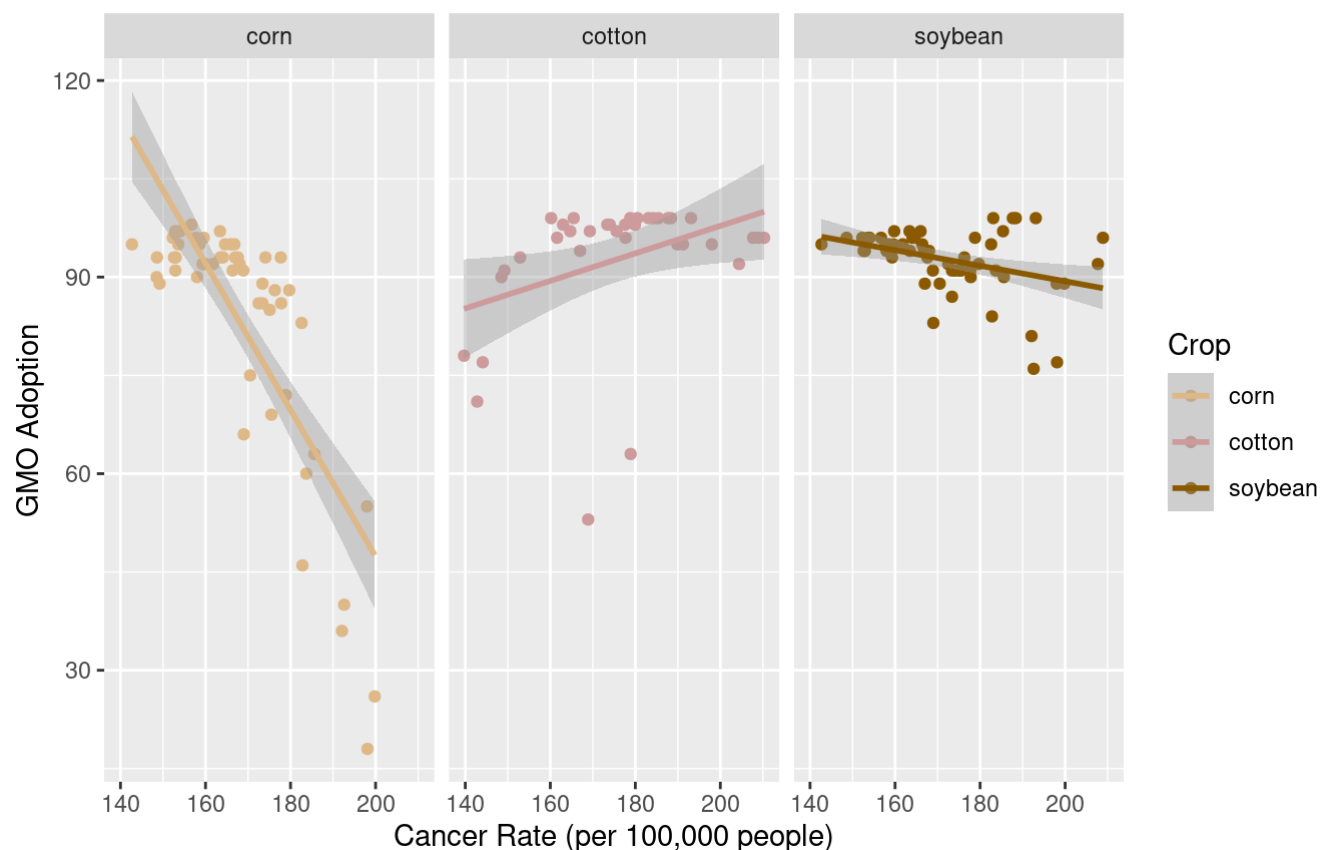
As pictured above a scatterplot was created to connect the relationship between the amount of cancer deaths, the GMP adoption amounts and its possible influence on the total policy number. In investigating this relationship, its clear that the higher the policy number pertaining to GMO adoption the lower the amount of deaths relating to cancer. In the visualization above, most of the data points, with a super high policy number and GMO adoption amount also have low cancer deaths, are clustered in the bottom right. In a similar manner we see a couple of high policy numbers pertaining to high cancer deaths at the top of the visualization.

```
# gmo and cancer rate by difference in crop
project2 %>%
  ggplot(aes(x = Rate, y = totalgmo, color = Crop)) +
  geom_point() +
  labs(title = 'Relationship between Cancer Incidence Rate and GMO Adoption by Crop (Fig 1.7)',
        subtitle = 'GMO vs. Cancer Incidence Rate vs. Crop Type ', x = 'Cancer Rate (per 100,000 people)', y = 'GMO Adoption')+
  facet_wrap(~Crop) +
  geom_smooth (method = 'lm') +
  scale_x_continuous(breaks=seq(0,300,20)) +
  scale_color_manual(values=c("burlywood", "rosybrown3", "orange4"))
```

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

## Relationship between Cancer Incidence Rate and GMO Adoption by Crop (Fig 1.8)

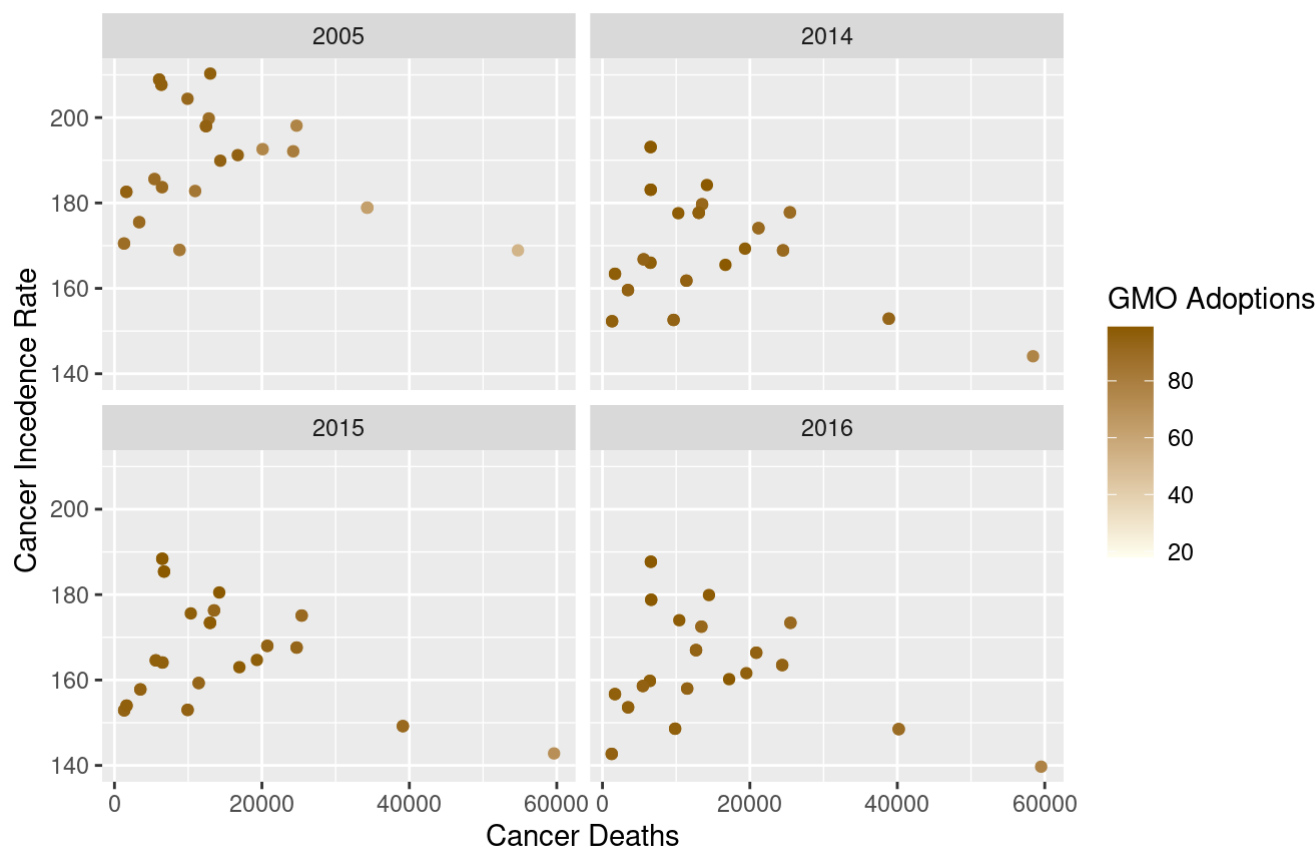
### GMO vs. Cancer Incidence Rate vs. Crop Type



A scatter plot was created in order to visualize the relationship between GMO amount, Crop Type and Cancer Rate. In doing so, it is plausible that some genetically engineered crops are related to an increase in cancer rate as they decrease such as that of soybean and corn. However, the visualization also displays that for cotton crops as the amount of production is increased, the cancer rate also increases.

```
# relationship between death rate and incidence rate with gmo
project2 %>%
  ggplot(aes(x = Deaths, y = Rate, color = totalgmo)) +
  geom_point() +
  labs(title = 'Relationship between Cancer Death Rate and Incidence Rate by GMO Value
(Fig 1.8)', subtitle = 'Cancer Deaths vs. Cancer Incidence Rate (per 100,000 people) v
s. Total GMO Adoption ', x = 'Cancer Deaths', y = 'Cancer Incidence Rate') +
  facet_wrap(~Year) +
  scale_color_continuous(low = "ivory", high = "orange4", name = "GMO Adoptions" )
```

## Relationship between Cancer Death Rate and Incidence Rate by GMO Value (Cancer Deaths vs. Cancer Incidence Rate (per 100,000 people) vs. Total GMO Adoption)



As pictured above a visualization was created to view the possible relationship between the amount of cancer deaths, the cancer rate and the total gmo adoption amounts for different given years. For instance, in 2005 there seems to be more higher gmo adoptions as indicated by the darker blue circles. In this same year we also see a higher Cancer incidence rate. On the other hand, in 2014, 2015, and 2016 there is a increase amount of gmo adoptions as signified by the light blue dots, however there is a wider spread in the distribution of Cancer Deaths.

## Discussion

### Do genetically engineered crops seem to have an influence on the cancer incidence rate and death rate in the United States?

Within our research research we hypothesized that the increase of GMO could correlate to an increase in cancer incidence and death rates. In Fig 1.8, visualization created to view the possible relationship between the amount of cancer deaths, the cancer rate and the total GMO adoption amounts for different given years, there seems to be a connection between all 3 variables. Within this visualization we see that as the years go by there is an increase in GMO adoptions and a constant range in cancer deaths over the years. Although the incidence rate decreases a bit in the latter years there is a more widespread distribution of deaths caused by Cancer. Within our visualization we see that the Cancer Incidence Rate used for this analysis report has a normal distribution with an average of 171.5 incidents of cancer per 100,000 people. This rate utilized can further explain why there is not much change in the incidence rate within our visualization throughout the years as the mean value within each state likely varied 16.06 standard deviations. Despite the clear negative correlation in Fig 1.6 between GMO Adoption and Cancer Deaths, the Fig 1.8 showing the relationship between cancer deaths and incidence displays that more GMO adoptions per year has some slight correlation to more cancerous deaths.

When taking a closer look at the potential differences by crop type in Fig 1.7, there are clear differences as cotton is the only crop with a positive correlation between incidence rate and GMO adoption. The other two have negative correlations with the corn being highly negative and soybean being slightly negative. Because cotton is used significantly less as a food product than the other two crop types this is a limitation of our study.

The relationships portrayed through the visualizations lead us to a conclusion that GMO adoption is not a significant predictor for cancerous deaths, but may have some impact on cancer incidence development.

### **What is the relationship between the number of food policy bills and the cancer incidence and death rates over years?**

Within this research we hypothesized that with an increase in food policy bills there would be a decrease in cancer deaths. In addition to this, the policy rates increasing are effective at protecting the quality of nutrients in the food market that could lower cancer rates. In Fig 1.6, a tri-variate graph represents the relationship between Cancer Death Rate, GMO Adoption and Total Policy Number, the higher the policy number pertaining to GMO adoption the lower the amount of deaths relating to cancer. In the visualization above, most of the data points, with a super high policy number and GMO adoption amount also have low cancer deaths, are clustered in the bottom right. In a similar manner we see a couple of high policy numbers pertaining to high cancer deaths at the top of the visualization with higher death rates, this could be related to legislation being created as a result of the high cancer deaths and not as a preventative measure. In further support of this relationship the skewed distribution of Deaths, the median deaths per 100,000 incidents to report is 11958 with the IQR reporting at 13641.

The scatterplot of the three variables and their relationships have visualized the inconsistent effect policy bills have on the impact of cancerous deaths despite the amount of GMO adoptions.

### **Overall Project Conduct**

Overall the process of the project was pretty smooth. The initial difficulty was first picking a research topic that would yield good findings. As Public Health Majors and fanatics of Data Science we decided to combine both passions and create a project that would allow us to be able to bridge both passions. As time moved on we decided to go off of our own speculation of GMOs to health effects, we wanted to explore its relationship to one type of chronic condition, cancer, that is often linked to nutrition and lifestyle factors. In also considering the effect of food and nutrition on the development of chronic conditions, we explored the how the possible regulations through legislation could have an the effect of food production quality which could impact cancer incidence. Within the process we have truly learned the importance of starting with clean and tidy data so that our project would make sense. As a result of this we had to adjust our project, but in the end it led to overall better visualizations and a project to be proud of.

*Acknowledgements: R Graph Gallery Visualization Graphing Website, Our Wonderful Professor and TA's. Overall the tidying, joining/merging and wrangling was done by Sedem K. and Visualization, and Discussion and formatting was done by Abena B. Oftentimes we would work on either section and discuss its improvements with one another.*