Project 1: Wrangling, Exploration, Visualization

SDS322E

Data Wrangling, Exploration, Visualization

Alex Kim UT Eid: ak37642

Introduction Paragraph or two introducing your datasets and variables, why they are interesting to you, etc.

```r
# read your datasets in here, e.g., with read_csv()
library(tidyverse)
library(readr)
library(fivethirtyeight)
data1 <- bad_drivers
data2 <- read_csv("~/project1/2021 County Health Rankings Data - v1.csv",
    skip = 1)
```

I chose the bad_drivers dataset from the fivethirtyeight package and the county health ratings from Countyhealthratings.org (under the Rankings data and document page). I wanted to look at these datasets because I am very big on keeping up with the news. One thing that I always see are news articles on drunk driving, speeding, or just not paying attention lead to people losing their lives due to these accidents. I know that these are just a few factors; therefore, I wanted to bring in the County health rankings data because the data provides more insight into the health aspect of these various counties in each state of the United States.

Some of the variables in my first dataset of "bad_drivers" include number of drivers involved in fatal collisions, percentage of those speeding, alcohol-impaired, not distracted, no prior accidents before this collision, cost of car insurance, and losses experienced by insurance companies across all 50 states. Furthermore, some of the variables for my second dataset of "County_Health_Rankings" include life expectancy, deaths, percent of ethnicities, number of people insured, and more across counties in the 50 states in the United States.I would expect to see that if there were more traffic accidents the location and how populated the state is, could impact the number of traffic fatalities. Furthermore, if there was an increase in the number of people uninsured, this could influence how the insurance premium will come out to be.

```r

# Joining the datasets
joineddata <- full_join(data1, data2, by = c(state = "State"))

# Finding which is missing from each dataset
not_in_data1 <- anti_join(data1, data2, by = c(state = "State"))
not_in_data2 <- anti_join(data2, data1, by = c(State = "state"))
```

Joining/Merging I combined both of my datasets by the common variable of "states." To do this, I utilized the full_join function to retain all of my rows, which is why there was 3,193 rows, one row for each county in each state. Furthermore, the two datasets were joined by the common ID of "state." When fully joined, there were 3,193 rows with 72 columns. Before the full join, the bad_drivers dataset had 51 rows of the 50 states in the U.S. as well as D.C. with 8 variables. Also before the join, the County health ratings data had 3,193 rows, for each county in the each state of the U.S. along with 65 variables. There were 51 unique

IDs, or states, for both the bad_drivers and county health rating datasets. Since both had only unique IDs of the 51 states, there were no IDs that were different from each other in both datasets. Since both datasets were joined by the common function of states, all 50 states in the United States including D.C. were included in both datasets. No observations were dropped and were actually retained by the full join function as well as the common ID being the same for each dataset. The NAs that are located in the full join and extra columns will be tidied. which will be shown above in the tidying section.

Tidying: Reshaping If your datasets are tidy already, demonstrate that you can reshape data with pivot wider/longer here (e.g., untidy and then retidy). Alternatively, it may be easier to wait until the wrangling section so you can reshape your summary statistics. Note here if you are going to do this.

```r
# your tidying code (if applicable; can also wait until
# wrangling section)
clean_joineddata <- joineddata %>% select(state, num_drivers,
    perc_alcohol, insurance_premiums, County, `Life Expectancy`,
    `# Deaths`, `% Frequent Physical Distress`, `# Uninsured`,
    `Median Household Income`, `Homicide Rate`, `# Asian`, `# American Indian & Alaska Native`,
    `# Native Hawaiian/Other Pacific Islander`, `# Hispanic`,
    `# Black`, `# Non-Hispanic White`, `% Rural`)

clean_joineddata <- clean_joineddata %>% rename(DriverDeaths = num_drivers,
    AlcoholRelated = perc_alcohol, LE = `Life Expectancy`, Deaths = `# Deaths`,
    PhysicalDistress = `% Frequent Physical Distress`, Uninsured = `# Uninsured`,
    HouseholdIncome = `Median Household Income`, Homicide = `Homicide Rate`,
    Asian = `# Asian`, AmericanIndianAlaskaNative = `# American Indian & Alaska Native`,
    PacificIslander = `# Native Hawaiian/Other Pacific Islander`,
    Hispanic = `# Hispanic`, Black = `# Black`, White = `# Non-Hispanic White`,
    PercRural = `% Rural`, Premiums = insurance_premiums)

clean_joineddata <- clean_joineddata %>% group_by(state) %>%
    summarize_at(c("LE", "Deaths", "DriverDeaths", "AlcoholRelated",
        "PhysicalDistress", "Uninsured", "HouseholdIncome", "Homicide",
        "PercRural", "Premiums", "Asian", "AmericanIndianAlaskaNative",
        "PacificIslander", "Hispanic", "Black", "White"), .funs = list(mean = mean))

clean_joineddata2 <- clean_joineddata %>% pivot_longer(-1) %>%
    separate(name, into = c("name", "stat")) %>% pivot_wider(names_from = "name",
    values_from = "value") %>% select(-stat) %>% mutate(Rural = ifelse(PercRural >
    50, "yes", "no"))
```

Since there were so many different variables, I wanted to just select 10 different unique variables and then variables of different ethnicities. I wanted to untidy and tidy my data first before I started on the wrangling portion, and I just wanted to clean up the names before moving on. In this case, I had to first select the columns I wanted and then remove the extraneous ones like percent of the population over 65, graduation rate, GPA, child mortality rate, infant mortality, etc. There were over 72 columns with the data joined. With this, I made a cleaner version of the joineddata (clean_joineddata). I then wanted to just make to create a new categorical variable of whether one state would be considered rural or not. Since the variable Rural was in percentages, I mutated the variable with the ifelse function to say that if the average 'PercRural' was greater than 50%, then we can say that it is considered rural. If not, then it wasn't considered rural. I then renamed all of the columns because I was having difficulty when it came to tidying the data with the special characters like %, _, and # signs. I then used the pivot_longer and pivot_wider functions to show that I know how to tidy. Since htere were so many repeats of states because of the various counties, I went ahead and actually made new columnns using pivot_longer and pivot_wider to reshape my data to show the averages across the states instead. With these, I found the means of each state with my numeric variables, including one where I mutated the function '% Rural' to give me a yes if the percentage of rural areas was greater than 50%, and no if otherwise. Through this, I was able to get better visualizations of my data because the plots looked really messy otherwise.

```
# Counting NAs
clean_joineddata2 %>% summarize_all(function(x) sum(is.na(x)))
```

Wrangling

```
## # A tibble: 1 x 18
##   state    LE Deaths DriverDeaths AlcoholRelated PhysicalDistress Uninsured
##   <int> <int>  <int>        <int>          <int>            <int>     <int>
## 1     0    17     14            0              0                0         1
## # ... with 11 more variables: HouseholdIncome <int>, Homicide <int>,
## #   PercRural <int>, Premiums <int>, Asian <int>,
## #   AmericanIndianAlaskaNative <int>, PacificIslander <int>, Hispanic <int>,
## #   Black <int>, White <int>, Rural <int>
```

When I used the summary function, I found that there was 17 NAs in Life Expectancy, 14 NAs in Deaths, 48 NAs in homicide rates, 2 NAs in percent of counties that are rural, 1 NA in number of people uninsured and median household income, 0 NAs in state, Driver fatalities, Alcohol related traffic fatalities, physical distress, insurance premiums, number of Asians, American Indian & Alaska Native, Pacific Islander, Hispanic, Black, and White.

```
# % of BIPOC in each county
clean_joineddata2 <- clean_joineddata2 %>% group_by(state) %>%
    mutate(`% BIPOC` = (sum(Asian, AmericanIndianAlaskaNative,
        PacificIslander, Hispanic, Black))/(sum(Asian, AmericanIndianAlaskaNative,
        PacificIslander, Hispanic, Black, White)) * 100)

clean_joineddata2 <- clean_joineddata2 %>% mutate(Diverse = ifelse(`% BIPOC` >
    50, "high", "low"))
```

First, I wanted to create a new variable that would tell us the percentage of many people of color make up the population in each state. To do this, I utilized the mutate function to sum up BIPOC populations, dividing it by the sum of the "BIPOC" populations and "White" populations, and then multiplying by 100. This gave me the percentage of BIPOC in each state within the U.S. I then utilized mutate again and the ifelse function to tell me if these states were considered diverse. If their percentages were greater than 50% then I considered it as high.

```
# Summary statistics for 10 numeric variables
clean_joineddata2 %>% summarize_at(c("DriverDeaths", "AlcoholRelated",
    "Premiums", "LE", "Deaths", "PhysicalDistress", "Uninsured",
    "HouseholdIncome", "Homicide", "PercRural", "Asian", "AmericanIndianAlaskaNative",
    "PacificIslander", "Hispanic", "Black", "White"), na.rm = TRUE,
    .funs = list(mean = mean, sd = sd, max = max, min = min,
        n = n_distinct)) %>% pivot_longer(contains("_")) %>%
    separate(name, into = c("variable", "stat")) %>% pivot_wider(names_from = "variable",
    values_from = "value") %>% knitr::kable()
```

state	stat	DriverDeaths	AlcoholRelated	Premiums	LE	Deaths	PhysicalDistress	Uninsured	HouseholdIncome	Homicide	PercRural	Asian	AmericanIndianAlaskaNative	PacificIslander	Hispanic	Black	White
Alabama	mean	18.8	30	784.57	74.69274	7469.085	15.72059	9062.44	46229.16	NaN	66.69256	8568.088	823.5000	150.2946	5267.00	8169.84	4192.00
Alabama	sd	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Alabama	max	18.8	30	784.57	74.69274	7469.085	15.72059	9062.44	46229.16	-Inf	66.69256	8568.088	823.5000	150.2946	5267.00	8169.84	4192.00

state	stat	Drivers	Alcohol	Premature	Infant	Deaths	Physical	Uninsured	Households	Homicides	Income	Asian	American Indian	Pacific Islander	Hispanic	Black	White
Alabama	min	18.8	30	784.57	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	Inf	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Alabama	n	1.0	1	1.00	1.0000	1.00000	1.00000	1.0000	1.00	0.0000	1.0000	1.0000	1.0000	1.00000	1.000	1.0000	1.00
Alaska	mean	18.1	25	1053.48	NaN	NaN	14.33333	[illegible]	[illegible]	NaN	NaN	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Alaska	sd	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Alaska	max	18.1	25	1053.48	-Inf	-Inf	14.33333	[illegible]	[illegible]	-Inf	-Inf	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Alaska	min	18.1	25	1053.48	Inf	Inf	14.33333	[illegible]	[illegible]	Inf	Inf	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Alaska	n	1.0	1	1.00	0.0000	0.00000	1.00000	1.0000	1.00	0.0000	0.0000	1.0000	1.0000	1.00000	1.000	1.0000	1.00
Arizona	mean	18.6	28	899.47	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	NaN	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Arizona	sd	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Arizona	max	18.6	28	899.47	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	-Inf	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Arizona	min	18.6	28	899.47	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	Inf	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Arizona	n	1.0	1	1.00	1.0000	1.00000	1.00000	1.0000	1.00	0.0000	1.0000	1.0000	1.0000	1.00000	1.000	1.0000	1.00
Arkansas	mean	22.4	26	827.37	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	NaN	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Arkansas	sd	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Arkansas	max	22.4	26	827.37	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	-Inf	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Arkansas	min	22.4	26	827.37	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	Inf	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Arkansas	n	1.0	1	1.00	1.0000	1.00000	1.00000	1.0000	1.00	0.0000	1.0000	1.0000	1.0000	1.00000	1.000	1.0000	1.00
California	mean	12.0	28	878.41	NaN	[illegible]	[illegible]	[illegible]	[illegible]	NaN	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
California	sd	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
California	max	12.0	28	878.41	-Inf	[illegible]	[illegible]	[illegible]	[illegible]	-Inf	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
California	min	12.0	28	878.41	Inf	[illegible]	[illegible]	[illegible]	[illegible]	Inf	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
California	n	1.0	1	1.00	0.0000	1.00000	1.00000	1.0000	1.00	0.0000	1.0000	1.0000	1.0000	1.00000	1.000	1.0000	1.00
Colorado	mean	13.6	28	835.50	NaN	NaN	11.15385	[illegible]	[illegible]	NaN	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Colorado	sd	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Colorado	max	13.6	28	835.50	-Inf	-Inf	11.15385	[illegible]	[illegible]	-Inf	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Colorado	min	13.6	28	835.50	Inf	Inf	11.15385	[illegible]	[illegible]	Inf	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Colorado	n	1.0	1	1.00	0.0000	0.00000	1.00000	1.0000	1.00	0.0000	1.0000	1.0000	1.0000	1.00000	1.000	1.0000	1.00
Connecticut	mean	10.8	36	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Connecticut	sd	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Connecticut	max	10.8	36	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Connecticut	min	10.8	36	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Connecticut	n	1.0	1	1.00	1.0000	1.00000	1.00000	1.0000	1.00	1.0000	1.0000	1.0000	1.0000	1.00000	1.000	1.0000	1.00
Delaware	mean	16.2	30	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Delaware	sd	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Delaware	max	16.2	30	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Delaware	min	16.2	30	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Delaware	n	1.0	1	1.00	1.0000	1.00000	1.00000	1.0000	1.00	1.0000	1.0000	1.0000	1.0000	1.00000	1.000	1.0000	1.00
District of Columbia	mean	5.9	27	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
District of Columbia	sd	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
District of Columbia	max	5.9	27	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

state	stat	Driver	Alcohol	Premature	Firearms	Deaths	Physical	Uninsured	Household	Homicide	Income	Asian	American Indian	Pacific Islander	Hispanic	Black	White
District of Columbia	min	5.9	27	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
District of Columbia	n	1.0	1	1.00	1.0000	1.0000	1.00000	1.0000	1.00	1.0000	1.0000	1.0000	1.0000	1.00000	1.000	1.0000	1.00
Florida	mean	17.9	29	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	NaN	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Florida	sd	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Florida	max	17.9	29	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	-Inf	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Florida	min	17.9	29	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	Inf	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Florida	n	1.0	1	1.00	1.0000	1.0000	1.00000	1.0000	1.00	0.0000	1.0000	1.0000	1.0000	1.00000	1.000	1.0000	1.00
Georgia	mean	15.6	25	913.15	NaN	[illegible]	[illegible]	[illegible]	[illegible]	NaN	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Georgia	sd	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Georgia	max	15.6	25	913.15	-Inf	[illegible]	[illegible]	[illegible]	[illegible]	-Inf	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Georgia	min	15.6	25	913.15	Inf	[illegible]	[illegible]	[illegible]	[illegible]	Inf	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Georgia	n	1.0	1	1.00	0.0000	1.0000	1.00000	1.0000	1.00	0.0000	1.0000	1.0000	1.0000	1.00000	1.000	1.0000	1.00
Hawaii	mean	17.5	41	861.18	NaN	NaN	10.83333	NaN	NaN	NaN	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Hawaii	sd	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Hawaii	max	17.5	41	861.18	-Inf	-Inf	10.83333	-Inf	-Inf	-Inf	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Hawaii	min	17.5	41	861.18	Inf	Inf	10.83333	Inf	Inf	Inf	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Hawaii	n	1.0	1	1.00	0.0000	0.0000	1.00000	0.0000	0.00	0.0000	1.0000	1.0000	1.0000	1.00000	1.000	1.0000	1.00
Idaho	mean	15.3	29	641.96	NaN	NaN	13.00000	[illegible]	[illegible]	NaN	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Idaho	sd	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Idaho	max	15.3	29	641.96	-Inf	-Inf	13.00000	[illegible]	[illegible]	-Inf	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Idaho	min	15.3	29	641.96	Inf	Inf	13.00000	[illegible]	[illegible]	Inf	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Idaho	n	1.0	1	1.00	0.0000	0.0000	1.00000	1.0000	1.00	0.0000	1.0000	1.0000	1.0000	1.00000	1.000	1.0000	1.00
Illinois	mean	12.8	34	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	NaN	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Illinois	sd	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Illinois	max	12.8	34	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	-Inf	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Illinois	min	12.8	34	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	Inf	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Illinois	n	1.0	1	1.00	1.0000	1.0000	1.00000	1.0000	1.00	0.0000	1.0000	1.0000	1.0000	1.00000	1.000	1.0000	1.00
Indiana	mean	14.5	29	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	NaN	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Indiana	sd	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Indiana	max	14.5	29	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	-Inf	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Indiana	min	14.5	29	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	Inf	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Indiana	n	1.0	1	1.00	1.0000	1.0000	1.00000	1.0000	1.00	0.0000	1.0000	1.0000	1.0000	1.00000	1.000	1.0000	1.00
Iowa	mean	15.7	25	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	NaN	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Iowa	sd	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Iowa	max	15.7	25	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	-Inf	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Iowa	min	15.7	25	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	Inf	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Iowa	n	1.0	1	1.00	1.0000	1.0000	1.00000	1.0000	1.00	0.0000	1.0000	1.0000	1.0000	1.00000	1.000	1.0000	1.00
Kansas	mean	17.8	24	780.45	NaN	NaN	[illegible]	[illegible]	[illegible]	NaN	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Kansas	sd	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Kansas	max	17.8	24	780.45	-Inf	-Inf	[illegible]	[illegible]	[illegible]	-Inf	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

state	stat	Drivers	Alcohol	Premiums	Infant	Deaths	Physical	Uninsured	Household	Homicide	Income	Asian	American Indian	Pacific Islander	Hispanic	Black	White
Kansas	min	17.8	24	780.45	Inf	Inf	12.09436	485.4155	51039.05	Inf	66.50179	1792.1661	131.5660	69.56604	4718.331	3157.0415	71450.25
Kansas	n	1.0	1	1.00	0.00000	0.00000	1.00000	1.0000	1.00	0.0000	1.00000	1.0000	1.0000	1.00000	1.000	1.0000	1.00
Kentucky	mean	21.4	23	872.51	74.63326	3228.7176	76.92479	639.8347	47490.40	NaN	71.40108	1081.8824	241.5537	69.28926	2887.760	6080.9752	62239.88
Kentucky	sd	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Kentucky	max	21.4	23	872.51	74.63326	3228.7176	76.92479	639.8347	47490.40	-Inf	71.40108	1081.8824	241.5537	69.28926	2887.760	6080.9752	62239.88
Kentucky	min	21.4	23	872.51	74.63326	3228.7176	76.92479	639.8347	47490.40	Inf	71.40108	1081.8824	241.5537	69.28926	2887.760	6080.9752	62239.88
Kentucky	n	1.0	1	1.00	1.00000	1.00000	1.00000	1.0000	1.00	0.0000	1.00000	1.0000	1.0000	1.00000	1.000	1.0000	1.00
Louisiana	mean	20.5	33	1281.75	55.54426	4218.4355	30.43077	7148.0388	38826.60	NaN	50.71239	2390.5232	1316.0308	87.81537	7599.136	36184.8335	33547.14
Louisiana	sd	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Louisiana	max	20.5	33	1281.75	55.54426	4218.4355	30.43077	7148.0388	38826.60	-Inf	50.71239	2390.5232	1316.0308	87.81537	7599.136	36184.8335	33547.14
Louisiana	min	20.5	33	1281.75	55.54426	4218.4355	30.43077	7148.0388	38826.60	Inf	50.71239	2390.5232	1316.0308	87.81537	7599.136	36184.8335	33547.14
Louisiana	n	1.0	1	1.00	1.00000	1.00000	1.00000	1.0000	1.00	0.0000	1.00000	1.0000	1.0000	1.00000	1.000	1.0000	1.00
Maine	mean	15.1	30	661.88	78.37261	2157.8223	23.94118	8715.5529	29490.18	NaN	73.43520	2049.4114	1149.0588	53.52941	1788.235	2535.7647	47011.41
Maine	sd	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Maine	max	15.1	30	661.88	78.37261	2157.8223	23.94118	8715.5529	29490.18	-Inf	73.43520	2049.4114	1149.0588	53.52941	1788.235	2535.7647	47011.41
Maine	min	15.1	30	661.88	78.37261	2157.8223	23.94118	8715.5529	29490.18	Inf	73.43520	2049.4114	1149.0588	53.52941	1788.235	2535.7647	47011.41
Maine	n	1.0	1	1.00	1.00000	1.00000	1.00000	1.0000	1.00	0.0000	1.00000	1.0000	1.0000	1.00000	1.000	1.0000	1.00
Maryland	mean	12.5	32	1048.78	78.51520	2601.8410	40.20000	3542.5769	73016.04	NaN	33.73325	2504.2940	2407.2000	536.40000	50505.760	48212.3200	136002.48
Maryland	sd	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Maryland	max	12.5	32	1048.78	78.51520	2601.8410	40.20000	3542.5769	73016.04	-Inf	33.73325	2504.2940	2407.2000	536.40000	50505.760	48212.3200	136002.48
Maryland	min	12.5	32	1048.78	78.51520	2601.8410	40.20000	3542.5769	73016.04	Inf	33.73325	2504.2940	2407.2000	536.40000	50505.760	48212.3200	136002.48
Maryland	n	1.0	1	1.00	1.00000	1.00000	1.00000	1.0000	1.00	0.0000	1.00000	1.0000	1.0000	1.00000	1.000	1.0000	1.00
Massachusetts	mean	8.2	35	1011.80	14.44667	6629.7333	13.13333	2260.1833	83279.07	NaN	16.16667	6335.4570	4570.6667	988.00000	13987.603	7604.2536	36740.00
Massachusetts	sd	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Massachusetts	max	8.2	35	1011.80	14.44667	6629.7333	13.13333	2260.1833	83279.07	-Inf	16.16667	6335.4570	4570.6667	988.00000	13987.603	7604.2536	36740.00
Massachusetts	min	8.2	35	1011.80	14.44667	6629.7333	13.13333	2260.1833	83279.07	Inf	16.16667	6335.4570	4570.6667	988.00000	13987.603	7604.2536	36740.00
Massachusetts	n	1.0	1	1.00	1.00000	1.00000	1.00000	1.0000	1.00	0.0000	1.00000	1.0000	1.0000	1.00000	1.000	1.0000	1.00
Michigan	mean	14.1	28	1110.68	71.10300	3048.1190	19.19048	8779.6936	53695.44	NaN	61.21890	1909.1429	1298.6905	99.66667	12576.321	3035.3877	71730.05
Michigan	sd	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Michigan	max	14.1	28	1110.68	71.10300	3048.1190	19.19048	8779.6936	53695.44	-Inf	61.21890	1909.1429	1298.6905	99.66667	12576.321	3035.3877	71730.05
Michigan	min	14.1	28	1110.68	71.10300	3048.1190	19.19048	8779.6936	53695.44	Inf	61.21890	1909.1429	1298.6905	99.66667	12576.321	3035.3877	71730.05
Michigan	n	1.0	1	1.00	1.00000	1.00000	1.00000	1.0000	1.00	0.0000	1.00000	1.0000	1.0000	1.00000	1.000	1.0000	1.00
Minnesota	mean	9.6	29	777.18	80.18016	1184.6591	59.17500	997.5964	69448.90	NaN	60.94564	5647.8876	1760.8864	95.77273	7162.046	4695.9318	18367.02
Minnesota	sd	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Minnesota	max	9.6	29	777.18	80.18016	1184.6591	59.17500	997.5964	69448.90	-Inf	60.94564	5647.8876	1760.8864	95.77273	7162.046	4695.9318	18367.02
Minnesota	min	9.6	29	777.18	80.18016	1184.6591	59.17500	997.5964	69448.90	Inf	60.94564	5647.8876	1760.8864	95.77273	7162.046	4695.9318	18367.02
Minnesota	n	1.0	1	1.00	1.00000	1.00000	1.00000	1.0000	1.00	0.0000	1.00000	1.0000	1.0000	1.00000	1.000	1.0000	1.00
Mississippi	mean	17.6	31	896.07	NaN	1242.6265	26.53494	813.0361	41753.66	NaN	70.38776	675.9545	130.7229	43.51807	2412.286	6834.4071	47439.33
Mississippi	sd	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Mississippi	max	17.6	31	896.07	-Inf	1242.6265	26.53494	813.0361	41753.66	-Inf	70.38776	675.9545	130.7229	43.51807	2412.286	6834.4071	47439.33
Mississippi	min	17.6	31	896.07	Inf	1242.6265	26.53494	813.0361	41753.66	Inf	70.38776	675.9545	130.7229	43.51807	2412.286	6834.4071	47439.33
Mississippi	n	1.0	1	1.00	0.00000	1.00000	1.00000	1.0000	1.00	0.0000	1.00000	1.0000	1.0000	1.00000	1.000	1.0000	1.00
Missouri	mean	16.1	34	790.37	26.76638	1381.3749	19.93103	1404.4693	65391.31	NaN	66.48229	3228.4612	627.9138	169.94828	2632.897	9270.8345	73410.21
Missouri	sd	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA

state	stat	Drivers	Alcohol	Premium	Firearms	Deaths	Physical Distress	Uninsured	Household Income	Homicide	Poor	Asian	American Indian	Pacific Islander	Alaska Native Hispanic	Black	White
Missouri	max	16.1	34	790.37	26.76653	1381.374	14.93103	3404.465	65591.31	-Inf	66.48292	3228.4627	627.9138	169.94846	2632.897	12270.8375	73410.21
Missouri	min	16.1	34	790.37	26.76653	1381.374	14.93103	3404.465	65591.31	Inf	66.48292	3228.4627	627.9138	169.94846	2632.897	12270.8375	73410.21
Missouri	n	1.0	1	1.00	1.00000	1.00000	1.00000	1.0000	1.00	0.0000	1.00000	1.00000	1.0000	1.00000	1.000	1.0000	1.00
Montana	mean	21.4	44	816.21	NaN	NaN	12.26316	510.4561	61629.21	NaN	74.57365	3465.0124	2594.5965	32.28070	1518.912	1022.3158	38200.39
Montana	sd	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Montana	max	21.4	44	816.21	-Inf	-Inf	12.26316	510.4561	61629.21	-Inf	74.57365	3465.0124	2594.5965	32.28070	1518.912	1022.3158	38200.39
Montana	min	21.4	44	816.21	Inf	Inf	12.26316	510.4561	61629.21	Inf	74.57365	3465.0124	2594.5965	32.28070	1518.912	1022.3158	38200.39
Montana	n	1.0	1	1.00	0.00000	0.00000	1.00000	1.0000	1.00	0.0000	1.00000	1.00000	1.0000	1.00000	1.000	1.0000	1.00
Nebraska	mean	14.9	35	732.28	NaN	NaN	10.55319	927.8723	56105.59	NaN	74.14163	1631.6633	323.0851	49.97872	2673.298	8817.6329	32695.15
Nebraska	sd	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Nebraska	max	14.9	35	732.28	-Inf	-Inf	10.55319	927.8723	56105.59	-Inf	74.14163	1631.6633	323.0851	49.97872	2673.298	8817.6329	32695.15
Nebraska	min	14.9	35	732.28	Inf	Inf	10.55319	927.8723	56105.59	Inf	74.14163	1631.6633	323.0851	49.97872	2673.298	8817.6329	32695.15
Nebraska	n	1.0	1	1.00	0.00000	0.00000	1.00000	1.0000	1.00	0.0000	1.00000	1.00000	1.0000	1.00000	1.000	1.0000	1.00
Nevada	mean	14.7	32	1029.87	NaN	NaN	14.00000	6360.664	64713.17	NaN	47.07292	2796.5272	1222.7778	2731.66667	10066.361	6879.5654	56681.44
Nevada	sd	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Nevada	max	14.7	32	1029.87	-Inf	-Inf	14.00000	6360.664	64713.17	-Inf	47.07292	2796.5272	1222.7778	2731.66667	10066.361	6879.5654	56681.44
Nevada	min	14.7	32	1029.87	Inf	Inf	14.00000	6360.664	64713.17	Inf	47.07292	2796.5272	1222.7778	2731.66667	10066.361	6879.5654	56681.44
Nevada	n	1.0	1	1.00	0.00000	0.00000	1.00000	1.0000	1.00	0.0000	1.00000	1.00000	1.0000	1.00000	1.000	1.0000	1.00
New Hampshire	mean	11.6	30	746.57	49.26386	2688.909	10.90909	1142.070	70350.00	NaN	53.94737	5327.2727	48.0000	122.18182	9225.273	3646.1818	18897.64
New Hampshire	sd	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
New Hampshire	max	11.6	30	746.57	49.26386	2688.909	10.90909	1142.070	70350.00	-Inf	53.94737	5327.2727	48.0000	122.18182	9225.273	3646.1818	18897.64
New Hampshire	min	11.6	30	746.57	49.26386	2688.909	10.90909	1142.070	70350.00	Inf	53.94737	5327.2727	48.0000	122.18182	9225.273	3646.1818	18897.64
New Hampshire	n	1.0	1	1.00	1.00000	1.00000	1.00000	1.0000	1.00	0.0000	1.00000	1.00000	1.0000	1.00000	1.000	1.0000	1.00
New Jersey	mean	11.2	28	1301.52	72.90845	5525.901	11.59090	6980.3636	85692.68	NaN	12.42857	26619.5989	3039.0000	930.90909	16880.4000	1000.4014	14909.55
New Jersey	sd	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
New Jersey	max	11.2	28	1301.52	72.90845	5525.901	11.59090	6980.3636	85692.68	-Inf	12.42857	26619.5989	3039.0000	930.90909	16880.4000	1000.4014	14909.55
New Jersey	min	11.2	28	1301.52	72.90845	5525.901	11.59090	6980.3636	85692.68	Inf	12.42857	26619.5989	3039.0000	930.90909	16880.4000	1000.4014	14909.55
New Jersey	n	1.0	1	1.00	1.00000	1.00000	1.00000	1.0000	1.00	0.0000	1.00000	1.00000	1.0000	1.00000	1.000	1.0000	1.00
New Mexico	mean	18.4	27	869.85	NaN	NaN	14.58824	4598.824	46532.91	NaN	47.67253	3238.8235	5517.2941	196.52941	10761.294	2238.4118	15448.82
New Mexico	sd	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
New Mexico	max	18.4	27	869.85	-Inf	-Inf	14.58824	4598.824	46532.91	-Inf	47.67253	3238.8235	5517.2941	196.52941	10761.294	2238.4118	15448.82
New Mexico	min	18.4	27	869.85	Inf	Inf	14.58824	4598.824	46532.91	Inf	47.67253	3238.8235	5517.2941	196.52941	10761.294	2238.4118	15448.82
New Mexico	n	1.0	1	1.00	0.00000	0.00000	1.00000	1.0000	1.00	0.0000	1.00000	1.00000	1.0000	1.00000	1.000	1.0000	1.00

state	stat	Driver	Alcohol	Premature	Infant	Deaths	Physical	Distress Uninsured	Household	Homicide	Income	Rural	Asian American	Indian Pacific Islander	Alaskan Hispanic	Native Black	White
New York	mean	12.3	29	1234.71	39.6760	9379.68	12.15873	3278.663	41962.84	NaN	43.7855	45657.60	1013.7460	858.2222	12908	18236.3	214041.90
New York	sd	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
New York	max	12.3	29	1234.71	39.6760	9379.68	12.15873	3278.663	41962.84	-Inf	43.7855	45657.60	1013.7460	858.2222	12908	18236.3	214041.90
New York	min	12.3	29	1234.71	39.6760	9379.68	12.15873	3278.663	41962.84	Inf	43.7855	45657.60	1013.7460	858.2222	12908	18236.3	214041.90
New York	n	1.0	1	1.00	1.0000	1.00000	1.00000	1.0000	1.00	0.0000	1.00000	1.0000	1.0000	1.00000	1.000	1.0000	1.00
North Carolina	mean	16.8	31	708.27	47.0846	4691.52	13.87322	7527.652	53698.40	NaN	60.9366	61689.73	3229.1485	260.0000	20313.4	14568.43	395041.62
North Carolina	sd	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
North Carolina	max	16.8	31	708.27	47.0846	4691.52	13.87322	7527.652	53698.40	-Inf	60.9366	61689.73	3229.1485	260.0000	20313.4	14568.43	395041.62
North Carolina	min	16.8	31	708.27	47.0846	4691.52	13.87322	7527.652	53698.40	Inf	60.9366	61689.73	3229.1485	260.0000	20313.4	14568.43	395041.62
North Carolina	n	1.0	1	1.00	1.0000	1.00000	1.00000	1.0000	1.00	0.0000	1.00000	1.0000	1.0000	1.00000	1.000	1.0000	1.00
North Dakota	mean	23.9	42	688.75	NaN	NaN	10.88889	429.2963	61698.94	NaN	81.7747	45963.29	1372.7778	22.77778	1167.852	520.9623	30611.59
North Dakota	sd	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
North Dakota	max	23.9	42	688.75	-Inf	-Inf	10.88889	429.2963	61698.94	-Inf	81.7747	45963.29	1372.7778	22.77778	1167.852	520.9623	30611.59
North Dakota	min	23.9	42	688.75	Inf	Inf	10.88889	429.2963	61698.94	Inf	81.7747	45963.29	1372.7778	22.77778	1167.852	520.9623	30611.59
North Dakota	n	1.0	1	1.00	0.0000	0.00000	1.00000	1.0000	1.00	0.0000	1.00000	1.0000	1.0000	1.00000	1.000	1.0000	1.00
Ohio	mean	14.1	34	697.73	36.8337	3719.93	13.52802	2877.652	52945.24	NaN	47.9363	57510.07	764.2697	158.5393	33572.33	18288.20	909047.01
Ohio	sd	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Ohio	max	14.1	34	697.73	36.8337	3719.93	13.52802	2877.652	52945.24	-Inf	47.9363	57510.07	764.2697	158.5393	33572.33	18288.20	909047.01
Ohio	min	14.1	34	697.73	36.8337	3719.93	13.52802	2877.652	52945.24	Inf	47.9363	57510.07	764.2697	158.5393	33572.33	18288.20	909047.01
Ohio	n	1.0	1	1.00	1.0000	1.00000	1.00000	1.0000	1.00	0.0000	1.00000	1.0000	1.0000	1.00000	1.000	1.0000	1.00
Oklahoma	mean	19.9	29	881.57	75.2810	5760.43	15.38462	2109.542	49800.10	NaN	63.2542	41316.79	3165.4872	220.0256	14233.75	5924.56	49158.67
Oklahoma	sd	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Oklahoma	max	19.9	29	881.57	75.2810	5760.43	15.38462	2109.542	49800.10	-Inf	63.2542	41316.79	3165.4872	220.0256	14233.75	5924.56	49158.67
Oklahoma	min	19.9	29	881.57	75.2810	5760.43	15.38462	2109.542	49800.10	Inf	63.2542	41316.79	3165.4872	220.0256	14233.75	5924.56	49158.67
Oklahoma	n	1.0	1	1.00	1.0000	1.00000	1.00000	1.0000	1.00	0.0000	1.00000	1.0000	1.0000	1.00000	1.000	1.0000	1.00
Oregon	mean	12.8	26	804.71	NaN	NaN	14.70270	1639.657	55768.57	NaN	43.6861	61966.91	1768.3784	1040.273	26740.34	3485.40	511107.24
Oregon	sd	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Oregon	max	12.8	26	804.71	-Inf	-Inf	14.70270	1639.657	55768.57	-Inf	43.6861	61966.91	1768.3784	1040.273	26740.34	3485.40	511107.24
Oregon	min	12.8	26	804.71	Inf	Inf	14.70270	1639.657	55768.57	Inf	43.6861	61966.91	1768.3784	1040.273	26740.34	3485.40	511107.24
Oregon	n	1.0	1	1.00	0.0000	0.00000	1.00000	1.0000	1.00	0.0000	1.00000	1.0000	1.0000	1.00000	1.000	1.0000	1.00
Pennsylvania	mean	18.2	31	905.97	78.2842	2366.47	10.72053	3539.951	51240.63	NaN	47.0514	41772.84	5929.7941	293.9412	29416.41	17618.52	591405.24
Pennsylvania	sd	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Pennsylvania	max	18.2	31	905.97	78.2842	2366.47	10.72053	3539.951	51240.63	-Inf	47.0514	41772.84	5929.7941	293.9412	29416.41	17618.52	591405.24

state	stat	Drivers	Alcohol	Premium	Losses	Deaths	Physical	Uninsured	Household	Homicide	Income	Asian	American Indian	Pacific Islander	Hispanic	Black	White
Pennsylvania	min	18.2	31	905.99	78.2842	2856.41	10.672053	3539.945	54240.63	Inf	47.0514	14172.85	529.7941	293.9412	29416.41	17618.52	289105.24
Pennsylvania	n	1.0	1	1.00	1.0000	1.00000	1.00000	1.0000	1.00	0.0000	1.0000	1.0000	1.0000	1.00000	1.000	1.0000	1.00
Rhode Island	mean	11.1	38	1148.89	99.4136	6874.333	13.366667	1489.3773	73302.67	NaN	11.1333	3378.3667	667.6667	708.3333	37548.00	20655.3333	233377.00
Rhode Island	sd	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Rhode Island	max	11.1	38	1148.89	99.4136	6874.333	13.366667	1489.3773	73302.67	-Inf	11.1333	3378.3667	667.6667	708.3333	37548.00	20655.3333	233377.00
Rhode Island	min	11.1	38	1148.89	99.4136	6874.333	13.366667	1489.3773	73302.67	Inf	11.1333	3378.3667	667.6667	708.3333	37548.00	20655.3333	233377.00
Rhode Island	n	1.0	1	1.00	1.0000	1.00000	1.00000	1.0000	1.00	0.0000	1.0000	1.0000	1.0000	1.00000	1.000	1.0000	1.00
South Carolina	mean	23.9	41	858.97	75.6136	3648.124	14.727660	2322.3494	49452.64	NaN	54.5240	4704.7224	231.5319	215.5319	31068.57	58186.1393	393469.87
South Carolina	sd	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
South Carolina	max	23.9	41	858.97	75.6136	3648.124	14.727660	2322.3494	49452.64	-Inf	54.5240	4704.7224	231.5319	215.5319	31068.57	58186.1393	393469.87
South Carolina	min	23.9	41	858.97	75.6136	3648.124	14.727660	2322.3494	49452.64	Inf	54.5240	4704.7224	231.5319	215.5319	31068.57	58186.1393	393469.87
South Carolina	n	1.0	1	1.00	1.0000	1.00000	1.00000	1.0000	1.00	0.0000	1.0000	1.0000	1.0000	1.00000	1.000	1.0000	1.00
South Dakota	mean	19.4	33	669.31	NaN	NaN	11.507463	392.0597	57070.91	NaN	NaN	408.9254	22488.1791	23.49254	1114.955	580.5075	71523.97
South Dakota	sd	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
South Dakota	max	19.4	33	669.31	-Inf	-Inf	11.507463	392.0597	57070.91	-Inf	-Inf	408.9254	22488.1791	23.49254	1114.955	580.5075	71523.97
South Dakota	min	19.4	33	669.31	Inf	Inf	11.507463	392.0597	57070.91	Inf	Inf	408.9254	22488.1791	23.49254	1114.955	580.5075	71523.97
South Dakota	n	1.0	1	1.00	0.0000	0.00000	1.00000	1.0000	1.00	0.0000	0.0000	1.0000	1.0000	1.00000	1.000	1.0000	1.00
Tennessee	mean	19.5	29	767.97	74.9442	4702.816	12.395833	1641.6500	50245.72	NaN	66.0928	2783.2580	580.4375	137.0418	6753.792	22787.29	194573.75
Tennessee	sd	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Tennessee	max	19.5	29	767.97	74.9442	4702.816	12.395833	1641.6500	50245.72	-Inf	66.0928	2783.2580	580.4375	137.0418	6753.792	22787.29	194573.75
Tennessee	min	19.5	29	767.97	74.9442	4702.816	12.395833	1641.6500	50245.72	Inf	66.0928	2783.2580	580.4375	137.0418	6753.792	22787.29	194573.75
Tennessee	n	1.0	1	1.00	1.0000	1.00000	1.00000	1.0000	1.00	0.0000	1.0000	1.0000	1.0000	1.00000	1.000	1.0000	1.00
Texas	mean	19.4	38	1004.75	NaN	NaN	13.98036	9708.254	56793.81	NaN	55.3631	11846.82	2215.9569	338.9179	95396.27	20463.93	376381.56
Texas	sd	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Texas	max	19.4	38	1004.75	-Inf	-Inf	13.98036	9708.254	56793.81	-Inf	55.3631	11846.82	2215.9569	338.9179	95396.27	20463.93	376381.56
Texas	min	19.4	38	1004.75	Inf	Inf	13.98036	9708.254	56793.81	Inf	55.3631	11846.82	2215.9569	338.9179	95396.27	20463.93	376381.56
Texas	n	1.0	1	1.00	0.0000	0.00000	1.00000	1.0000	1.00	0.0000	1.0000	1.0000	1.0000	1.00000	1.000	1.0000	1.00
Utah	mean	11.3	16	809.38	NaN	NaN	12.10000	4633.3667	63396.87	NaN	47.0557	5705.1333	3314.6667	2265.2667	30803.45	2087.0667	166250.60
Utah	sd	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Utah	max	11.3	16	809.38	-Inf	-Inf	12.10000	4633.3667	63396.87	-Inf	47.0557	5705.1333	3314.6667	2265.2667	30803.45	2087.0667	166250.60
Utah	min	11.3	16	809.38	Inf	Inf	12.10000	4633.3667	63396.87	Inf	47.0557	5705.1333	3314.6667	2265.2667	30803.45	2087.0667	166250.60
Utah	n	1.0	1	1.00	0.0000	0.00000	1.00000	1.0000	1.00	0.0000	1.0000	1.0000	1.0000	1.00000	1.000	1.0000	1.00
Vermont	mean	13.6	30	716.20	79.4399	2987.733	13.666667	748.1333	63085.87	NaN	74.4133	1335.2800	320.4667	32.9333	1695.867	1086.9733	373305.20
Vermont	sd	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA

state	stat	[illegible]	[illegible]	[illegible]	[illegible]	Deaths	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	White
Vermont	max	13.6	30	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	-Inf	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Vermont	min	13.6	30	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	Inf	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Vermont	n	1.0	1	1.00	1.0000	1.0000	1.00000	1.0000	1.00	0.0000	1.0000	1.0000	1.0000	1.00000	1.000	1.0000	1.00
Virginia	mean	12.7	27	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	NaN	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Virginia	sd	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Virginia	max	12.7	27	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	-Inf	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Virginia	min	12.7	27	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	Inf	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Virginia	n	1.0	1	1.00	1.0000	1.0000	1.00000	1.0000	1.00	0.0000	1.0000	1.0000	1.0000	1.00000	1.000	1.0000	1.00
Washington	mean	10.6	33	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	NaN	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Washington	sd	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Washington	max	10.6	33	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	-Inf	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Washington	min	10.6	33	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	Inf	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Washington	n	1.0	1	1.00	1.0000	1.0000	1.00000	1.0000	1.00	0.0000	1.0000	1.0000	1.0000	1.00000	1.000	1.0000	1.00
West Virginia	mean	23.8	28	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	NaN	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
West Virginia	sd	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
West Virginia	max	23.8	28	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	-Inf	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
West Virginia	min	23.8	28	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	Inf	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
West Virginia	n	1.0	1	1.00	1.0000	1.0000	1.00000	1.0000	1.00	0.0000	1.0000	1.0000	1.0000	1.00000	1.000	1.0000	1.00
Wisconsin	mean	13.8	33	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	NaN	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Wisconsin	sd	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Wisconsin	max	13.8	33	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	-Inf	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Wisconsin	min	13.8	33	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	Inf	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Wisconsin	n	1.0	1	1.00	1.0000	1.0000	1.00000	1.0000	1.00	0.0000	1.0000	1.0000	1.0000	1.00000	1.000	1.0000	1.00
Wyoming	mean	17.4	32	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	NaN	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Wyoming	sd	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Wyoming	max	17.4	32	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	-Inf	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Wyoming	min	17.4	32	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	Inf	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Wyoming	n	1.0	1	1.00	1.0000	1.0000	1.00000	1.0000	1.00	0.0000	1.0000	1.0000	1.0000	1.00000	1.000	1.0000	1.00

```r
clean_joineddata2 %>% summary(is.numeric)
```

```
##     state                 LE            Deaths         DriverDeaths  
##  Length:51          Min.   :74.63   Min.   :  587.7   Min.   : 5.90  
##  Class :character   1st Qu.:76.78   1st Qu.: 1501.4   1st Qu.:12.75  
##  Mode  :character   Median :78.32   Median : 2691.5   Median :15.60  
##                     Mean   :77.91   Mean   : 3793.2   Mean   :15.79  
##                     3rd Qu.:79.36   3rd Qu.: 5671.8   3rd Qu.:18.50  
##  AlcoholRelated  PhysicalDistress   Uninsured       HouseholdIncome
##  Min.   :16.00   Min.   : 9.00    Min.   :  248.1   Min.   :41754  
##  1st Qu.:28.00   1st Qu.:11.59    1st Qu.: 1013.8   1st Qu.:51647  
##  Median :30.00   Median :12.99    Median : 1678.6   Median :58552  
##  Mean   :30.69   Mean   :13.11    Mean   : 2677.5   Mean   :59654  
```

```
##  3rd Qu.:33.00   3rd Qu.:14.57    3rd Qu.: 2804.0   3rd Qu.:64647
##     Homicide        PercRural       Premiums          Asian
##  Min.   : 2.333   Min.   : 0.00   Min.   : 642.0   Min.   :   345
##  1st Qu.: 3.792   1st Qu.:43.78   1st Qu.: 768.4   1st Qu.:  1901
##  Median : 5.250   Median :54.16   Median : 859.0   Median :  5800
##  Mean   : 8.528   Mean   :51.04   Mean   : 887.0   Mean   : 19619
##  3rd Qu.:11.625   3rd Qu.:64.60   3rd Qu.:1007.9   3rd Qu.: 16595
##  AmericanIndianAlaskaNative PacificIslander       Hispanic
##  Min.   :  163.8            Min.   :   18.93   Min.   :  1113
##  1st Qu.:  756.1            1st Qu.:   94.72   1st Qu.:  5556
##  Median : 1760.9            Median :  175.29   Median : 12454
##  Mean   : 3917.9            Mean   : 1540.42   Mean   : 48860
##  3rd Qu.: 3973.8            3rd Qu.:  716.53   3rd Qu.: 50897
##      Black              White           Rural              % BIPOC
##  Min.   :   202.3   Min.   : 21524   Length:51          Min.   : 5.512
##  1st Qu.:  2847.1   1st Qu.: 63476   Class :character   1st Qu.:19.323
##  Median : 15211.2   Median :113526   Mode  :character   Median :27.949
##  Mean   : 34948.7   Mean   :166065                      Mean   :31.110
##  3rd Qu.: 41505.8   3rd Qu.:231980                      3rd Qu.:41.914
##    Diverse
##  Length:51
##  Class :character
##  Mode  :character
##
##
##  [ reached getOption("max.print") -- omitted 2 rows ]
```

```r
clean_joineddata2 %>% group_by(state) %>% summarize(count = n())
```

```
## # A tibble: 51 x 2
##    state                count
##    <chr>                <int>
##  1 Alabama                  1
##  2 Alaska                   1
##  3 Arizona                  1
##  4 Arkansas                 1
##  5 California               1
##  6 Colorado                 1
##  7 Connecticut              1
##  8 Delaware                 1
##  9 District of Columbia     1
## 10 Florida                  1
## # ... with 41 more rows
```

```r
clean_joineddata2 %>% group_by(Rural) %>% summarize(count = n())
```

```
## # A tibble: 3 x 2
##   Rural count
##   <chr> <int>
## 1 no       20
## 2 yes      29
## 3 <NA>      2
```

```r
clean_joineddata2 %>% group_by(Diverse) %>% summarize(count = n())
```

```
## # A tibble: 2 x 2
```

```
##   Diverse count
##   <chr>   <int>
## 1 high        6
## 2 low        45
```

This is a table of my numeric variables, which shows the mean, standard deviation, maximum, minimum, and distinct counts for each state in the United States. This is one big table that prints out these values for the states. I then found the counts for each of my categorical variables. As expected, there is only one count for each of the states. Surprisingly, the amount of states that were considered to be rural were almost even (with 20 being no and 31 being yes). Furthermore, very surprising to see that there were only 6 states that were considered to have high diversity (having a percentage of BIPOC greater than 50).

```
# Some more summary statistics for my variables
clean_joineddata2 %>% group_by(state, Diverse) %>% summarize(HouseholdIncome = mean(HouseholdIncome)) %
    arrange(desc((HouseholdIncome)))
```

```
## # A tibble: 51 x 3
## # Groups:   state [51]
##    state                Diverse HouseholdIncome
##    <chr>                <chr>             <dbl>
##  1 District of Columbia high             90395
##  2 New Jersey           low              85193.
##  3 Massachusetts        low              80279.
##  4 Maryland             low              79316.
##  5 Connecticut          low              79151.
##  6 Rhode Island         low              77003.
##  7 California           high             71197.
##  8 New Hampshire        low              70350
##  9 Alaska               low              67924.
## 10 Delaware             low              67517.
## # ... with 41 more rows
```

I wanted to compute some extra summary statistics for some of my variables. I grouped by state and Diverse, summarized to find the mean household incomes within each state, and then arranged by descending income. It was very interesting to see that the mean income in D.C. was $90395.00 and the diversity was considered to be high. The lowest average household income was in Mississippi with $41753.00 and the diversity level was considered low.

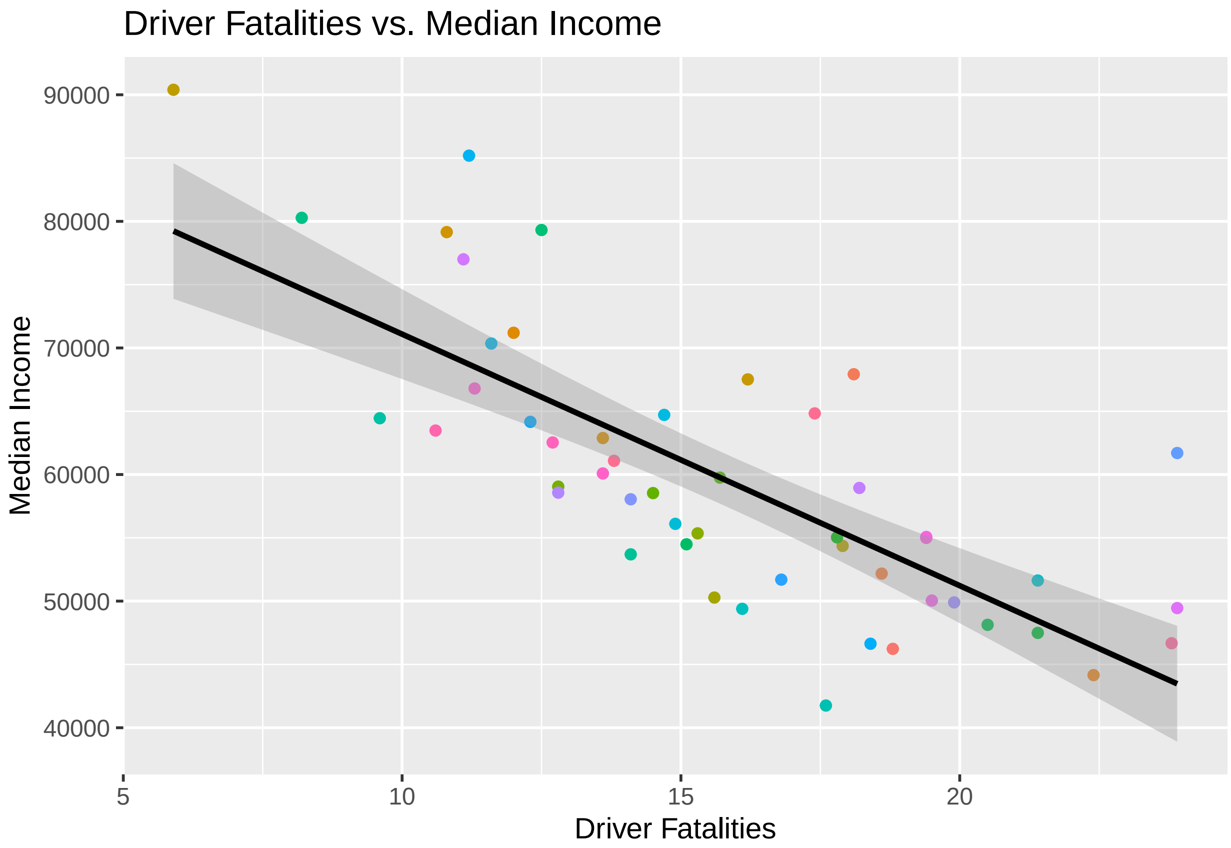
```
clean_joineddata2 %>% filter(Rural == "yes") %>% filter(str_detect(state,
    "^[aeiouAEIOU]"))
```

```
## # A tibble: 7 x 20
## # Groups:   state [7]
##   state    LE Deaths DriverDeaths AlcoholRelated PhysicalDistress Uninsured
##   <chr> <dbl>  <dbl>        <dbl>          <dbl>            <dbl>     <dbl>
## 1 Alab~  74.7  2419.         18.8             30             15.7     1062.
## 2 Arka~  75.0  1270.         22.4             26             16.6      828.
## 3 Idaho  NA      NA          15.3             29             13       1248.
## 4 Illi~  78.0  2801.         12.8             34             12.5     1927.
## 5 Indi~  77.0  1983.         14.5             29             13.2     2287.
## 6 Iowa   79.1   716.         15.7             25             10.2      408.
## 7 Okla~  75.3  1560.         19.9             29             15.4     2110.
## # ... with 13 more variables: HouseholdIncome <dbl>, Homicide <dbl>,
## #   PercRural <dbl>, Premiums <dbl>, Asian <dbl>,
## #   AmericanIndianAlaskaNative <dbl>, PacificIslander <dbl>, Hispanic <dbl>,
## #   Black <dbl>, White <dbl>, Rural <chr>, `% BIPOC` <dbl>, Diverse <chr>
```

Lastly, I wanted to filter and find the states that were considered to be rural. This gave me a value of 31, which meant out of the 50 states, we could say that around 31 were considered to be rural with an average percentage of their counties being greater than 50%. I then used the str_detect function to find the states out of the 31 that started with a vowel. This led me to Alabama, Alaska, Arkansas, Colorado, Georgia, Idaho, Illinois, Indiana, Iowa, and Kansas.

```r
# Plot of driver fatalities vs. median household income
clean_joineddata2 %>% ggplot(aes(x = DriverDeaths, y = HouseholdIncome)) +
    geom_point(aes(color = state)) + geom_smooth(method = "lm",
    color = 1) + scale_x_continuous(breaks = seq(0, 25, 5)) +
    scale_y_continuous(breaks = seq(0, 1e+05, 10000)) + ggtitle("Driver Fatalities vs. Median Income")
    xlab("Driver Fatalities") + ylab("Median Income") + theme(legend.position = "none")
```

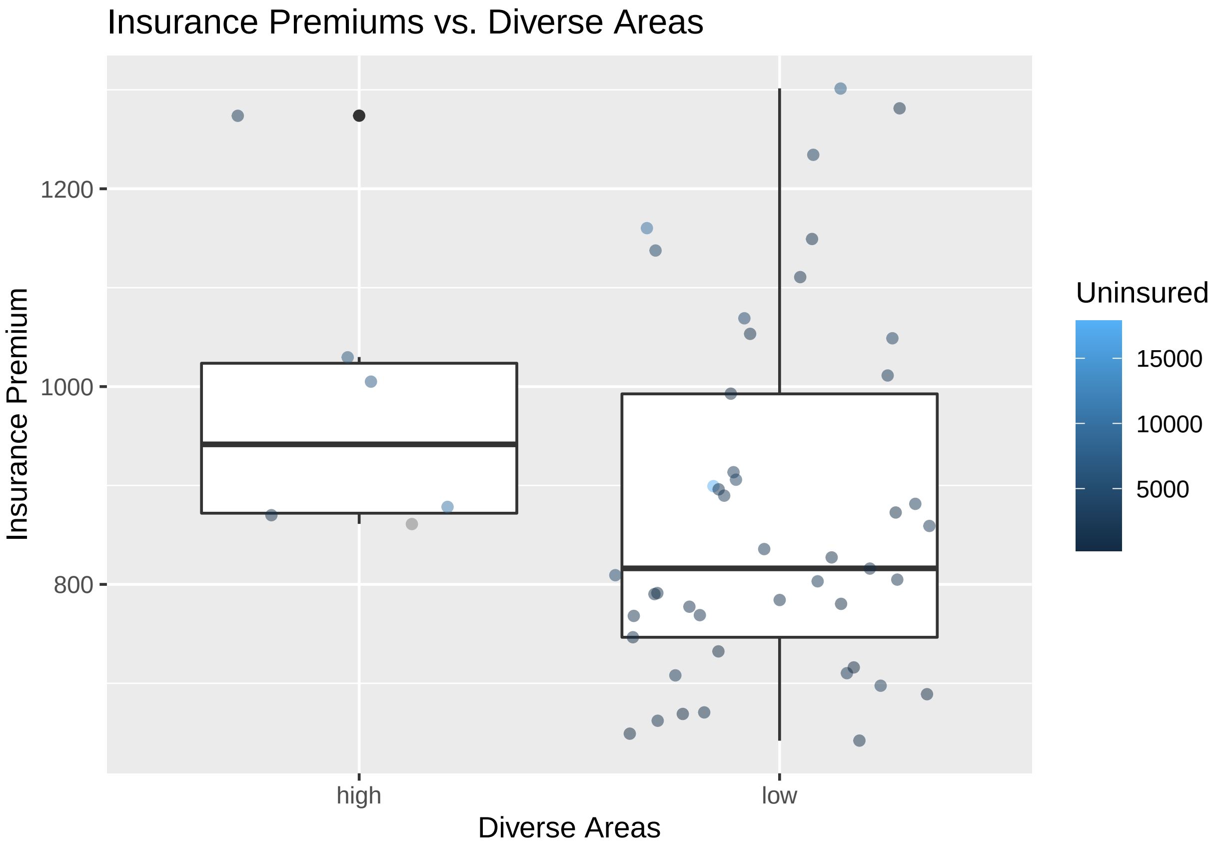
Visualizing



For the first ggplot, we're looking at the relationship between the number of drivers that have died from a car accident and median household income in each of the 50 U.S. states. Furthermore, i added points that are colored by state to help visualize the differences in each state for driver fatalities and median income. When looking at this plot, we can tell that there seems to be a relatively strong relationship between median household income and driver fatalities. As the median household income in each state decreased, there was a greater number of traffic fatalities that occurred. This could be from various different factors, but I thought it was interesting to see this possible relationship.

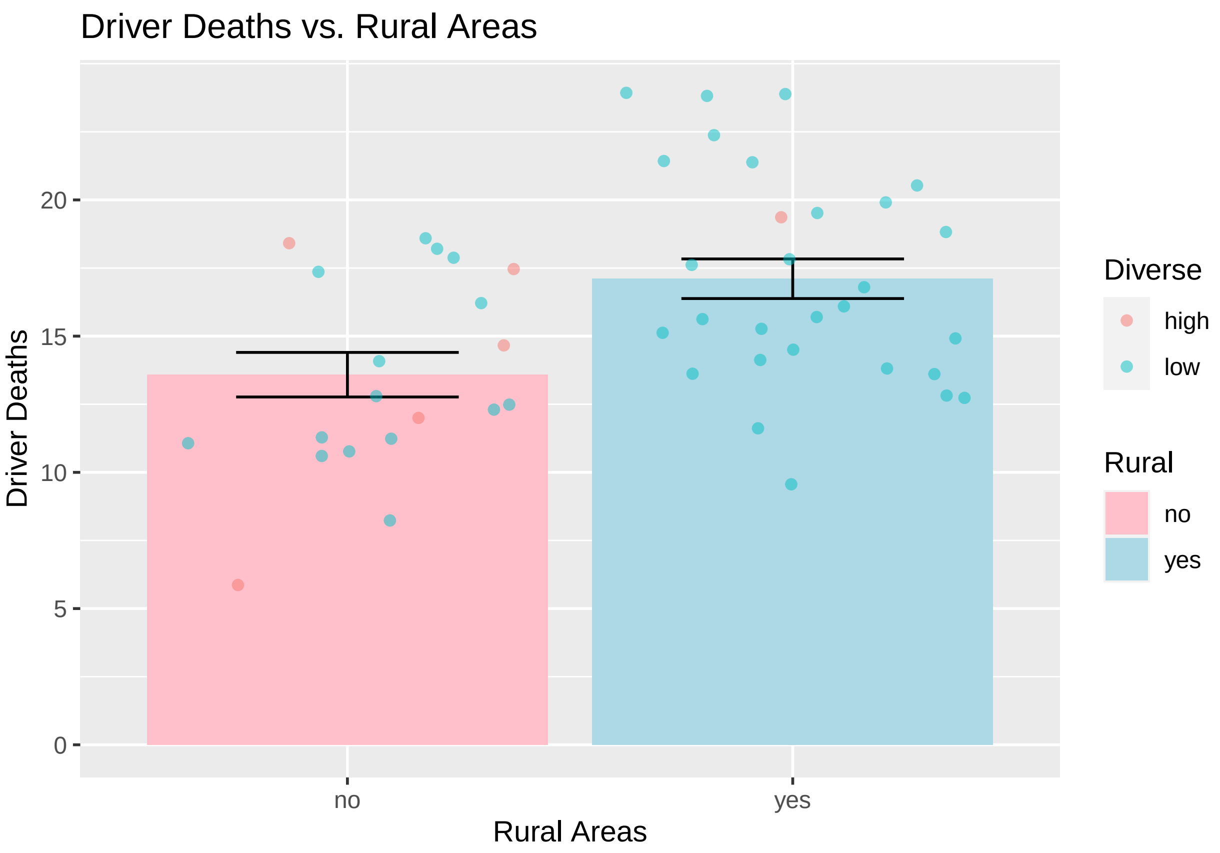
```r
# Plot of driver fatalities in each state in the United
# States
clean_joineddata2 %>% ggplot(aes(x = Diverse, y = Premiums)) +
```

```
    geom_boxplot() + scale_y_continuous(breaks = seq(0, 1400,
    200)) + ggtitle("Insurance Premiums vs. Diverse Areas") +
    xlab("Diverse Areas") + ylab("Insurance Premium") + geom_jitter(alpha = 0.5,
    aes(color = Uninsured))
```



For the next plot, I wanted to do a boxplot looking at the relationship between the diversity of the areas and the costs of insurance premiums across the state. I also wanted to overly the points with number of people uninsured to look at how these are all related to each other. There seems to be a positive skew for this plot. In states where there was low diversity, there seemed to be a range of various values of people uninsured and lower median insurance premium than those with high diversity. This could be due to a variety of reasons. In this case, we are only comparing the diversity areas based off of the percentages calculated in the wrangling section; therefore, this could be different if we looked at it from a county point of view. Just as in other boxplots, this plot gives us insight into the visualization of the minimum, first quartile, median, third quartile, and maximum, essentially the summary of the distribution.

```
# your plot 3
clean_joineddata2 %>% filter(!is.na(Rural)) %>% ggplot(aes(x = Rural,
    y = DriverDeaths)) + geom_bar(aes(fill = Rural), stat = "summary",
    fun = mean) + ggtitle("Driver Deaths vs. Rural Areas") +
    geom_errorbar(stat = "summary", width = 0.5) + xlab("Rural Areas") +
    ylab("Driver Deaths") + scale_fill_manual(values = c("pink",
    "light blue")) + scale_y_continuous(breaks = seq(0, 20, 5)) +
    geom_jitter(alpha = 0.5, aes(color = Diverse))
```



For my last plot, I wanted to take a look at the number of driver deaths in rural areas with an overlay of Diversity on the points. This was very interesting because we see that in cases of rural states, there is very lower diversity. Both rural and not rural have relatively similar error bars. In this case, rural areas seemed to have a higher amount of driver fatalities than states that weren't considered to be rural. In cases that aren't rural, there is still more states that are not as diverse. This could also be because states in the United States in general aren't super diverse.

Concluding Remarks I really enjoyed looking at these datasets. Keeping up with the news and learning about these incidents in various ways is informing. Before looking at these datasets, I really only focused on data that I saw locally and maybe a couple of nationwide data. I hope that I can continue to be conscious of the ways that I think about the news and how I view them across the states as well. Furthermore, I moved my tidying section to be after the joining becuase my knitting process wouldn't work otherwise.