

# Class Project

Eunice Olorunshola

11/21/2021

6760 CSC Big data programming

# Hypothesis

**“It is better to be safe than sorry”**

# Datasets Used with Links:

**Diabetes deaths per state**

**01**

\*([https://www.cdc.gov/nchs/pressroom/sosmap/diabetes\\_mortality/diabetes.htm](https://www.cdc.gov/nchs/pressroom/sosmap/diabetes_mortality/diabetes.htm))

**Deaths caused by distracted drivers, and alcohol impaired drivers.**

**03**

\*(<https://data.cdc.gov/Motor-Vehicle/Occupant-and-Alcohol-Impaired-Driving-Deaths-in-St/haed-k2ka>)

**Traffic Violations**

**02**

\*(<https://www.kaggle.com/felix4guti/traffic-violations-in-usa>)

**Personal Income**

**04**

\*(<https://www.kaggle.com/trinaghosh346/personal-income-expenditure>)

# Data Formatting

Diabetes deaths per state dataset not cleaned

CSV				
YEAR	STATE	RATE	DEATHS	URL
2019	AL	19.8	1224	/nchs/pressroom/states/alabama/al.htm
2019	AK	16.2	111	/nchs/pressroom/states/alaska/ak.htm
2019	AZ	23.4	2173	/nchs/pressroom/states/arizona/az.htm
2019	AR	29.8	1144	/nchs/pressroom/states/arkansas/ar.htm
2019	CA	21.8	9854	/nchs/pressroom/states/california/ca.htm
2019	CO	16.4	1046	/nchs/pressroom/states/colorado/co.htm
2019	CT	15.6	761	/nchs/pressroom/states/connecticut/ct.htm
2019	DE	20.5	267	/nchs/pressroom/states/delaware/de.htm
2019	FL	19.3	6174	/nchs/pressroom/states/florida/fl.htm
2019	GA	20.5	2375	/nchs/pressroom/states/georgia/ga.htm
2019	HI	15.8	311	/nchs/pressroom/states/hawaii/hi.htm
2019	ID	20.2	422	/nchs/pressroom/states/idaho/id.htm
2019	IL	18.1	2825	/nchs/pressroom/states/illinois/il.htm
2019	IN	25	2045	/nchs/pressroom/states/indiana/in.htm
2019	IA	20.2	861	/nchs/pressroom/states/iowa/ia.htm
2019	KS	24.4	884	/nchs/pressroom/states/kansas/ks.htm
2019	KY	29.1	1611	/nchs/pressroom/states/kentucky/ky.htm
2019	LA	26.5	1458	/nchs/pressroom/states/louisiana/la.htm
2019	ME	23.6	477	/nchs/pressroom/states/maine/me.htm
2019	MD	20.5	1514	/nchs/pressroom/states/maryland/md.htm
2019	MA	15.3	1383	/nchs/pressroom/states/massachusetts/ma.htm
2019	MI	22.2	2916	/nchs/pressroom/states/michigan/mi.htm
2019	MN	19.8	1404	/nchs/pressroom/states/minnesota/mn.htm
2019	MS	32.5	1159	/nchs/pressroom/states/mississippi/ms.htm
		20.9	1658	/nchs/pressroom/states/misc.htm

Diabetes deaths per state dataset cleaned

diabetescleaned				
	YEAR	STATE	RATE	DEATHS
301	2005	AK	22.4	93
251	2014	AK	19.4	113
201	2015	AK	23.9	142
151	2016	AK	19.3	124
101	2017	AK	19.3	130
51	2018	AK	17.7	122
1	2019	AK	16.2	111
300	2005	AL	29.8	1429
250	2014	AL	22.8	1281
200	2015	AL	21.7	1255
150	2016	AL	20.1	1183
100	2017	AL	19.8	1173
50	2018	AL	19.4	1176
0	2019	AL	19.8	1224
303	2005	AR	27.1	824
253	2014	AR	24.0	828
203	2015	AR	24.7	886
153	2016	AR	25.4	920
103	2017	AR	32.4	1180
53	2018	AR	32.3	1194
3	2019	AR	29.8	1144
302	2005	AZ	20.4	1208
252	2014	AZ	24.3	1936
202	2015	AZ	25.3	2081
152	2016	AZ	23.9	2025

# Data Formatting CONTD

Personal Income  
dataset not cleaned

Personal Income					
Posting Date	Day of the Week	Amount	Category	Description	Source
8/3/18	Friday	-1360	Bills	Rent	Chase Debit Card
8/4/18	Saturday	-117.89	Groceries		Chase Debit Card
8/4/18	Saturday	-34.27	Merchandise	Makeup/Skin Care	Chase Debit Card
8/4/18	Saturday	-21.59	Merchandise	Amazon	Chase Debit Card
8/5/18	Sunday	-10.79	Merchandise	Amazon	Chase Debit Card
8/10/18	Friday	-30.78	Merchandise	Amazon	Chase Debit Card
8/11/18	Saturday	-34.76	Merchandise	Clothes	Chase Debit Card
8/12/18	Sunday	-12.54	Merchandise	Amazon	Chase Debit Card
8/15/18	Wednesday	-9.98	Merchandise	Amazon	Chase Debit Card
8/16/18	Thursday	2500	Wire Transfer		Chase Debit Card
8/24/18	Friday	-70	Travel		Chase Debit Card
8/29/18	Wednesday	-75.06	Merchandise	Makeup/Skin Care	Chase Debit Card
8/31/18	Friday	-107	Groceries		Chase Debit Card
9/4/18	Tuesday	-1360	Bills	Rent	Chase Debit Card
9/10/18	Monday	-122.52	Groceries		Chase Debit Card
9/10/18	Monday	-50	Groceries		Chase Debit Card
9/19/18	Wednesday	-21.6	Merchandise	Best Buy	Discover Credit Card
9/22/18	Saturday	-16.19	Merchandise	Best Buy	Discover Credit Card
9/24/18	Monday	-46.5	Groceries		Chase Debit Card
9/28/18	Friday	-84.77	Groceries		Chase Debit Card
10/1/18	Monday	-1380	Bills	Rent	Chase Debit Card
10/1/18	Monday	-9.1	Pharmacy		Chase Debit Card
10/1/18	Monday	-7.72	Food		Chase Debit Card
	Monday	113.54	Wire Transfer		Chase Debit Card
		-55.1	Merchandise	Clothes	Chase Debit Card
		-22.28	Merchandise	Clothes	Chase Debit Card

Personal Income  
dataset cleaned

Personal_Income_datasetscleaned						
	Posting Date	Day of the Week	Amount Spent	Category	Description	Payment Method
205	1/1/19	Tuesday	16.2	Merchandise	Cosmetics	Discover Credit Card
236	1/11/19	Friday	44.4	Bills	Phone	Chase Debit Card
238	1/11/19	Friday	13.99	Merchandise	Amazon	Chase Debit Card
239	1/13/19	Sunday	24.98	Merchandise	Amazon	Discover Credit Card
240	1/14/19	Monday	372.59	Merchandise	Best Buy	Discover Credit Card
244	1/15/19	Tuesday	7.99	Merchandise	Amazon	Discover Credit Card
245	1/16/19	Wednesday	20.6	Merchandise	Amazon	Discover Credit Card
206	1/2/19	Wednesday	72.05	Merchandise	Clothes	Discover Credit Card
207	1/2/19	Wednesday	45.83	Merchandise	Cosmetics	Discover Credit Card
208	1/2/19	Wednesday	43.58	Merchandise	Cosmetics	Discover Credit Card
209	1/2/19	Wednesday	11.74	Merchandise	Amazon	Discover Credit Card
210	1/2/19	Wednesday	9.35	Merchandise	Clothes	Discover Credit Card
249	1/20/19	Sunday	102.59	Merchandise	Best Buy	Discover Credit Card
252	1/22/19	Tuesday	72.08	Merchandise	Clothes	Chase Debit Card
255	1/23/19	Wednesday	14.42	Merchandise	Amazon	Discover Credit Card
259	1/24/19	Thursday	9.95	Merchandise	Amazon	Chase Debit Card
261	1/25/19	Friday	11.74	Merchandise	Amazon	Discover Credit Card
263	1/28/19	Monday	7.88	Merchandise	Amazon	Discover Credit Card
267	1/29/19	Tuesday	14.77	Bills	Credit Card	Discover Credit Card
212	1/3/19	Thursday	31.11	Merchandise	Amazon	Discover Credit Card
213	1/3/19	Thursday	9.99	Merchandise	Amazon	Discover Credit Card
220	1/7/19	Monday	10.79	Merchandise	Amazon	Discover Credit Card
221	1/8/19	Tuesday	117.27	Merchandise	Cosmetics	Discover Credit Card
223	1/8/19	Tuesday	52.52	Merchandise	Clothes	Discover Credit Card
	1/8/19	Tuesday	28.05	Merchandise	Clothes	Discover Credit Card



# Data Formatting CONTD

Deaths caused by  
distracted and alcohol  
impaired drivers  
Dataset not cleaned

Occupant\_and\_Alcohol-

State	Alcohol-Impaired Driving Deaths	Occupant Deaths	Location
WY	484	1195	WY (43.23558,
NC	3996	10093	NC (35.47995,
OR	1129	2735	OR (44.567912
NY	3605	7013	NY (42.826892
AL	2997	7754	AL (32.839182
RI	265	429	RI (41.707447
WI	2300	4572	WI (44.395526
PA	4355	9656	PA (40.789676
NJ	1747	3832	NJ (40.104655
VA	2518	6335	VA (37.542699
TX	13171	24013	TX (31.827233
DE	430	767	DE (39.008676
MA	1309	2465	MA (42.268546
	968	1962	NV (39.491837
	631		

Deaths caused by distracted and alcohol impaired drivers dataset cleaned

occupant\_alcohol-

State	Alcohol-Impaired Driving Deaths	Occupant Deaths
41 AK	205	439
4 AL	2997	7754
26 AR	1636	4526
50 AZ	2656	5712
43 CA	9791	20733
44 CO	1599	3521
32 CT	1049	1806
29 DC	96	130
11 DE	430	767
27 FL	8053	16041
30 GA	3513	10406
25 HI	453	587
16 IA	967	2970
35 ID	661	1725
48 IL	3554	7348
31 IN	2171	5966
38 KS	1142	3248
21 KY	1900	6133
40 LA	2875	6073
12 MA	1309	2465
24 MD	1586	3454
23 ME	454	1231
18 MI	2710	6699
28 MN	1274	3248
46 MO	3004	7219
	2367	6100

# Data Formatting CONTD

Traffic Violations  
dataset not cleaned

Traffic Violations dataset  
cleaned

Sheet 1

Date Of Stop	Time Of Stop	Agency	SubAgency	Description
09/24/2013	17:11:00	MCP	3rd district, Silver Spring	DRIVING VEHICLE ON HIGHWAY WITH SUSPENDED REGISTRATION
12/20/2012	00:41:00	MCP	2nd district, Bethesda	DRIVING WHILE IMPAIRED BY ALCOHOL
07/20/2012	23:12:00	MCP	5th district, Germantown	FAILURE TO STOP AT STOP SIGN
03/19/2012	16:10:00	MCP	2nd district, Bethesda	DRIVER USING HANDS TO USE HANDHELD TELEPHONE WHILE MOTOR VEHICLE IN MOTION
12/01/2014	12:52:00	MCP	6th district, Gaithersburg / Montgomery Village	FAILURE TO STOP AND YIELD AT THRU HWY
06/09/2012	21:19:00	MCP	3rd district, Silver Spring	OCCUPANT UNDER 16 NOT RESTRAINED BY SEATBELT
09/11/2012	21:47:00	MCP	4th district, Wheaton	PERSON DRIVING MOTOR VEHICLE ON HIGHWAY OR PUBLIC USE PROPERTY
11/04/2012	01:14:00	MCP	4th district, Wheaton	DRIVING UNDER THE INFLUENCE OF ALCOHOL PER SE
08/27/2012	00:04:00	MCP	5th district, Germantown	PERSON DRIVING MOTOR VEHICLE WHILE LICENSE SUSPENDED UNDER 17-1
05/21/2012	16:47:00	MCP	2nd district, Bethesda	OPERATOR NOT RESTRAINED BY SEATBELT
08/27/2013	00:55:00	MCP	2nd district, Bethesda	NEGLIGENT DRIVING VEHICLE IN CARELESS AND IMPRUDENT MANNER ENDA
10/08/2013	13:23:00	MCP	4th district, Wheaton	DRIVING VEHICLE ON HIGHWAY WITH SUSPENDED REGISTRATION
04/24/2015	00:38:00	MCP	1st district, Rockville	DRIVER FAIL TO STOP AT FLASHING RED TRAFFIC SIGNAL STOP LINE
01/12/2014	20:56:00	MCP	1st district, Rockville	PERSON DRIVING MOTOR VEHICLE WHILE LICENSE SUSPENDED UNDER 17-1
02/14/2014	20:10:00	MCP	1st district, Rockville	FAILURE TO DRIVE ON RIGHT HAND ROADWAY OF DIVIDED HWY
04/09/2014	22:22:00	MCP	3rd district, Silver Spring	DRIVING VEHICLE ON HIGHWAY WITH AN EXPIRED LICENSE
12/06/2012	12:34:00	MCP	Headquarters and Special Operations	PEDESTRIAN CROSSING ROADWAY BETWEEN ADJACENT INTERSECTIONS H/
12/07/2012	18:36:00	MCP	1st district, Rockville	DRIVING VEHICLE WHILE UNDER THE INFLUENCE OF ALCOHOL
09/22/2012	21:24:00	MCP	3rd district, Silver Spring	DRIVER FAILURE TO OBEY PROPERLY PLACED TRAFFIC CONTROL DEVICE IN/
07/23/2013	23:15:00	MCP	4th district, Wheaton	FAILURE TO ATTACH VEHICLE REGISTRATION PLATES AT FRONT AND REAR
10/28/2013	17:41:00	MCP	6th district, Gaithersburg / Montgomery Village	DRIVER FAILURE TO YIELD RIGHT OF WAY ENTERING HWY. FROM PRIVATE RD
10/27/2012	19:13:00	MCP	3rd district, Silver Spring	FAILURE TO STOP AT STOP SIGN
02/11/2014	17:38:00	MCP	4th district, Wheaton	DRIVING VEHICLE ON HIGHWAY WITH SUSPENDED REGISTRATION
12/17/2015	13:15:00	MCP	5th district, Germantown	FAILURE TO RETURN TO & REMAIN AT SCENE OF ACCIDENT INVOLVING ATTEN
03/20/2013	08:53:00	MCP	4th district, Wheaton	EXCEEDING MAXIMUM SPEED: 49 MPH IN A POSTED 40 MPH ZONE
05/12/2015	08:26:00	MCP	1st district, Rockville	PERSON DRIVING MOTOR VEHICLE ON HIGHWAY OR PUBLIC USE P
	07:47:00	MCP	1st district, Rockville	PERSON DRIVING MOTOR VEHICLE WHILE LICENSE SUSPENDED

Sheet 1

TrafficViolations\_datasetscleaned

Location	Accident	Belts	Personal Injury	Property Damage	Fatal	Alcohol	State	Year	Violation Type	Contributed To Acc.
WOODMONT AVE @ RUGBY AVE	No	No	No	No	No	No	VA	2008	Warning	No
WOODMONT AVE @ MONTGOMERY LANE	No	No	No	No	No	No	VA	2008	Warning	No
WOODMONT AVE @ BATTERY LN	No	No	No	No	No	No	VA	2012	Warning	No
WOODFIELD RD AND VALLEY PARK DR	No	No	No	No	No	No	MD	2006	Citation	No
WOODFIELD RD AND VALLEY PARK DR	No	No	No	No	No	No	MD	2006	Warning	No
WISTERIA AT WALTER JOHNSON DR	No	No	No	No	No	No	MD	2006	Warning	No
WISTERIA @ FATHER HURLEY	No	No	No	No	No	No	MD	2008	Warning	No
WISCONSIN AVE/ MONTGOMERY AVE	No	No	No	No	No	No	MD	2004	Warning	No
WISCONSIN AVE/ MONTGOMERY AVE	No	No	No	No	No	No	MD	2004	Citation	No
WISCONSIN AVE AT MONTGOMERY	No	No	No	No	No	No	MD	2011	Warning	No
WISCONSIN AVE AT BETHESDA AVE	No	No	No	No	No	No	MD	2005	Warning	No
WISCONSIN AVE AT BETHESDA AVE	No	No	No	No	No	No	MD	2005	Warning	No
WISCONSIN AVE AT BETHESDA AVE	No	No	No	No	No	No	MD	2005	Warning	No
WISCONSIN AVE AT BETHESDA AVE	No	No	No	No	No	No	MD	2005	Warning	No
WISCONSIN AVE @ MAPLE AVE	No	No	No	No	No	No	MD	2007	Warning	No
WISCONSIN / OLD GEORGETOWN	No	No	No	No	No	No	MD	1998	Citation	No
WAVERLY ST @ WISCONSIN AVE	No	No	No	No	No	No	MD	2006	Warning	No
WAVERLY ST @ WISCONSIN AVE	No	No	No	No	No	No	MD	2006	Citation	No
WASHINGTONIAN BLVD @ SAM EIG HWY	No	No	No	No	No	No	MD	1996	Warning	No
WARING STATION RD / WISTERIA DR	No	No	No	No	No	No	MD	2014	Warning	No
W/B MIDDLEBROOK RD @ I-270	No	No	No	No	No	No	OH	2011	Warning	No
W/B MIDDLEBROOK RD @ I-270	No	No	No	No	No	No	OH	2011	Warning	No
W DIAMOND AVE @ PERRY PKWY	No	No	No	No	No	No	MD	1999	Warning	No
W DIAMOND AVE / MUDDY BRANCH	No	No	No	No	No	No	MD	2012	Warning	No
W DIAMOND AVE / MUDDY BRANCH	No	No	No	No	No	No	MD	2012	Warning	No
W DIAMOND AVE / MUDDY BRANCH	No	No	No	No	No	No	MD	2012	Warning	No



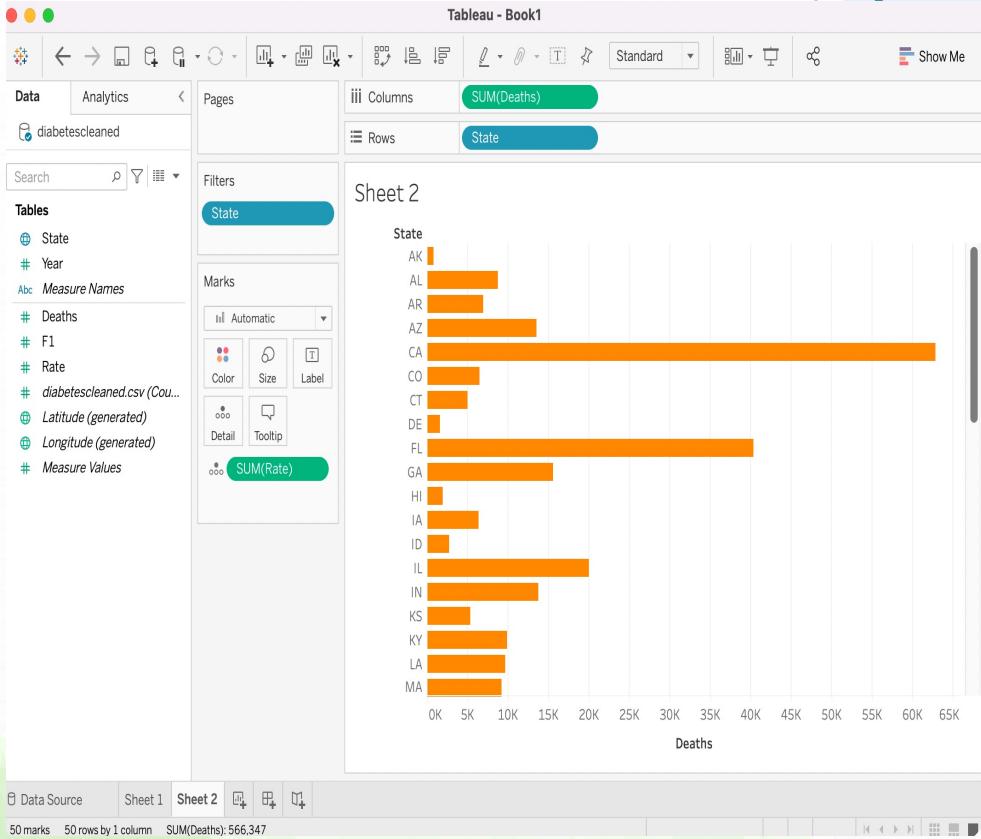
# How was the datasets cleaned?

- Deleted columns from the datasets that are not needed.
  - Deleted duplicates from the datasets.
- Renamed some of the columns to make it look more simple.
  - Changed some of the datatypes
  - Removed characters that were not needed.
  - Removed all rows with NaN in the columns.
  - Sort some of the columns in order.

# Tableau Visualization

## 01 Diabetes deaths per state dataset

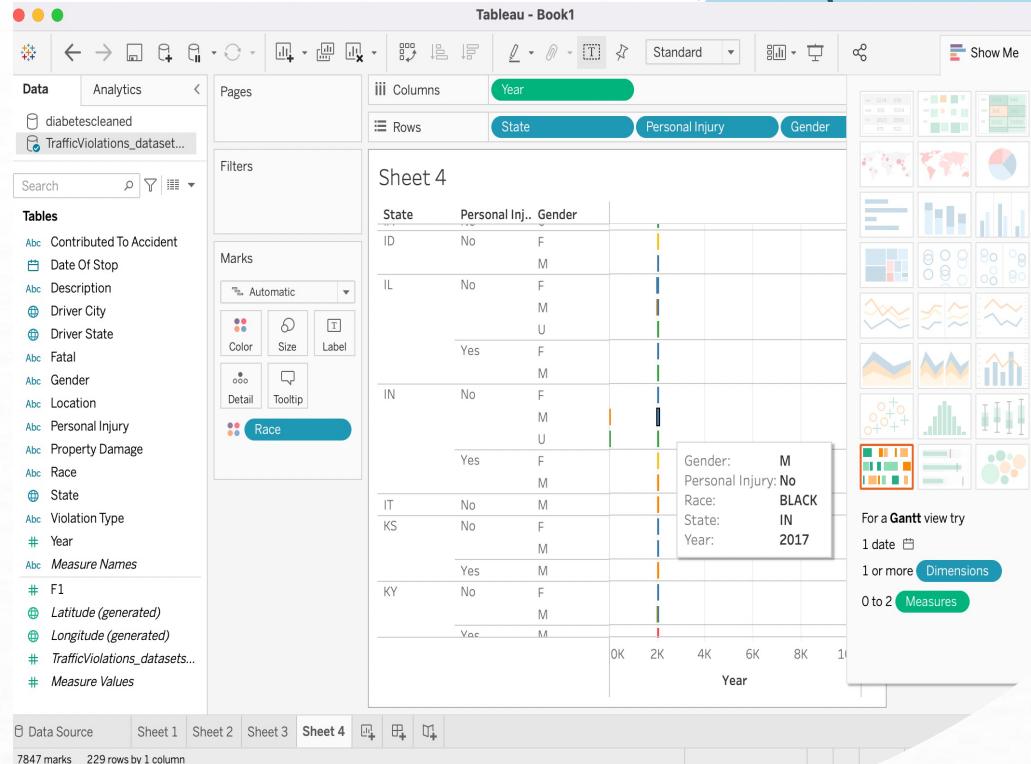
The picture to the right is a tableau visualization of a horizontal bar chart that shows diabetes deaths per state according to the dataset.



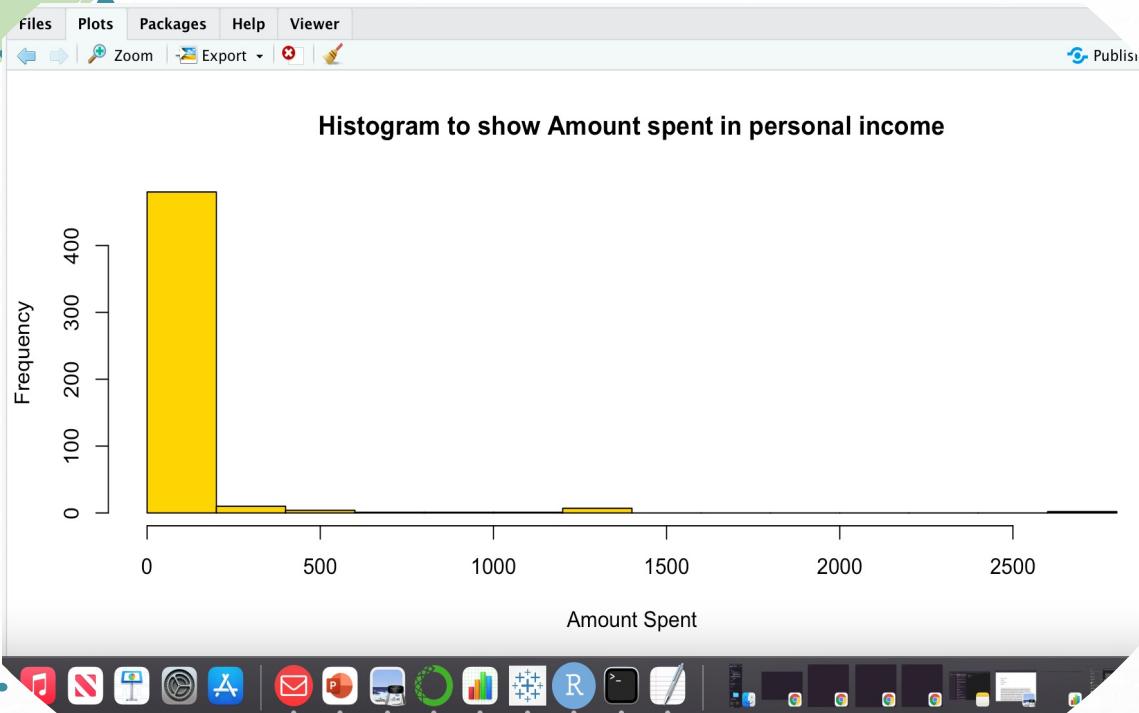
# Tableau Visualization CONT'D

## Traffic Violations dataset

- The picture to the right is a tableau visualization that shows the traffic violations dataset in a Ganatt view.
- I added state, personal injury, and gender in rows and year in columns.
- I did it this way to show per state and year by gender and race. Also, if there was an injury that was caused by not obeying traffic laws.



# R visualization



**Personal Income dataset**

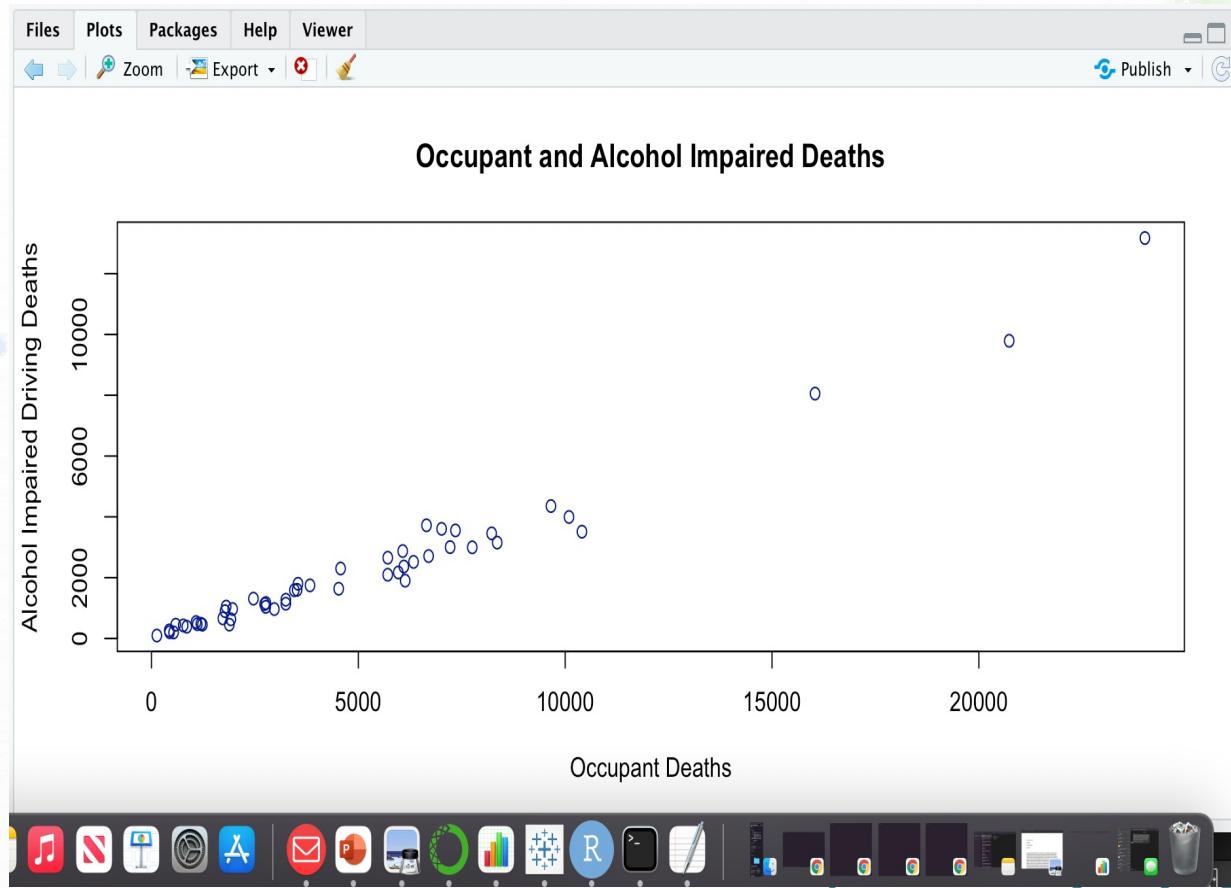
The picture to the left is a histogram created in R studio to show how there income was used each day of the week.

# R Visualization CONTD

01

**Deaths caused by  
distracted and alcohol  
impaired drivers dataset.**

- The picture to the right is a scatter plot shown in R studio
- The scatter plot shows occupant and alcohol impaired driving deaths per state.



# Tools/Technologies Used



## CSV

Saved the datasets as csv files. And used the files to create visualizations in Tableau and R studio.

R

## R Studio

- Used as a visualization
- Created a scatter plot for the deaths caused by distracted and alcohol impaired drivers dataset.
- Created a histogram for the personal income dataset.



## Tableau Desktop

- Used as a visualization for dataset diabetes deaths per state and traffic violations.
- The visualizations are horizontal bar chart and Ganatt view.

# Conclusion

- My hypothesis “it is better to be safe than sorry” has been proven to be true.
- From the datasets that I used each showed why it is always better to be safe than sorry.
- The first dataset diabetes deaths per state proves it by showing why it is important to take diabetes seriously and always follow up on medication and keeping track of sugar levels than to be sorry when it is not properly treated and lead to death or more harm to the body.
- The second dataset personal income proves it by showing why it is important to always save money and have money set aside just in case an emergency happens that you are not prepared for.
- The third dataset deaths caused by distracted and alcohol impaired drivers proves it by showing how it is better and smarter to always be focus on the road while driving and never drink and drive, because it can lead to death or someone else death that can put you in prison for life.
- The last dataset traffic violations also proves my hypothesis by showing why it is always important to follow the rule of the law while driving instead of disobeying it and getting arrested or putting other people lives in danger that can ultimately lead to death or injuries that can be severe.