Analyzing the NEISS Data

Moneka Bommasani

Introduction:

This is an analysis report on NEISS data.

Source:

United States Consumer Product Safety Commission's National Electronic Injury Surveillance System (NEISS) - http://www.cpsc.gov/en/Research--Statistics/NEISS-Injury-Data/

Tools Used:

RStudio	For Q1 – Q4
JMP	For Q5

Questions:

First, let's load the data into RStudio using "read.csv".
Now import the required libraries:

- library(dplyr)

1)

a) What are the top three body parts most frequently represented in this dataset?

Code:

- bodyparts <- left_join(NEISS2014, BodyParts, by = c("body_part" = "Code"))
- tail(names(sort(table(bodyparts$BodyPart))),3)

Result:

"Finger" "Face" "Head"

Explanation:

First, let's join the two datasets (NEISS2014, BodyParts), so as to display the name of the body parts. Now we can calculate the frequency of each body part with the table() function; then sort them to know the top three body parts using tail() function. This is because, by default sort gives an ASCENDING order. So we need the least 3 values to get the top three frequent body parts.

b) What are the top three body parts that are least frequently represented?

Code:

- bodyparts <- left_join(NEISS2014, BodyParts, by = c("body_part" = "Code"))
- head(names(sort(table(bodyparts$BodyPart))),3)

Result:

"25-50% of body"   "Pubic region"   "Not Recorded"

Explanation:

Same as above, we join the two data sets to display the name of the body part. Now we can calculate the frequency of each body part with the table() function; then sort them to know the least three body parts using head() function. This is because, by default sort gives an ASCENDING order. So we need the top 3 values to get the bottom three frequent body parts.

Note:

If we need the exact body parts, then the answer will be as follows:

- head(names(sort(table(bodyparts$BodyPart))),5)

"Pubic region”   "Internal"   "Arm, upper"

2)

a) How many injuries in this dataset involve a skateboard?
b) Of those injuries, what percentage were male and what percentage were female?
c) What was the average age of someone injured in an incident involving a skateboard?

<u>Approach 1:</u>

Code:

- nrow(data.frame(grep("skateboard", NEISS2014$narrative, ignore.case = TRUE)))

- subset <- NEISS2014[grep("skateboard", NEISS2014$narrative, ignore.case = TRUE),]
- options(digits=4)
- transform(as.data.frame(table(subset$sex)),percentage_column=Freq/nrow(subset)*100)

- +(subset$age > 200)
- mean(subset$age)
- summary(subset$age)

Result:

```
466

Var1          Freq   percentage_column
1 Female      82        17.6
2 Male        384       82.4

17.99
```

Explanation:

We can search the narrative section of the data for the term “skateboard” using grep() function and count the number of rows to find out the number of injuries involving skateboard.

Then, we can subset the above data to find out the number and percentage of males and females involved in skateboard injuries.

Now, let’s check if there are any kids less than 2 years involved in the skateboard injuries. The output shows that there are none. So, the mean or summary of the age column can be seen to find out the average age of a person injured in the skateboard incident.

<u>Approach 2:</u>

If we see the manual, the code for skateboard is 1333. Using the product code:

Code:

- skate <- subset(NEISS2014 , NEISS2014$prod1 == 1333 | NEISS2014$prod2 == 1333)
- nrow(skate)

- transform(as.data.frame(table(skate$sex)),percentage_column=Freq/nrow(skate)*100)

- +(subset$age > 200)
- mean(skate$age)
- summary(skate$age)

Result:

```
495

Var1          Freq   percentage_column
1 Female      89             17.98
2 Male 406           82.02

17.89
```

Explanation:
We can now use the product code 1333 (using logical OR for prod1 and prod2) to get the subset of data involving skateboard injuries and count the number of rows.
Then, we can calculate the frequency and percentage of males and females involved in skateboard injuries.
Now, let's check if there are any kids less than 2 years involved in the skateboard injuries. The output shows that there are none. So, the mean or summary of the age column can be seen to find out the average age of a person injured in the skateboard incident.

Note:

There is a difference in the results for both the approaches. This is because of the misspelling of the word "skateboard" in the narrative section of the original dataset in some instances.
So, the **approach 2** result is more appropriate.

Also, the manual includes "skateboard" injuries under some other products (5042, 5035, 3215, 5036, 5044) which are not considered in this result.

3)

a) What diagnosis had the highest hospitalization rate?

Code:

- Diagno <- left_join(NEISS2014, DiagnosisCodes, by = c("diag" = "Code"))
- hosp <- subset(Diagno , Diagno$disposition == 4)
- tail(names(sort(table(hosp$Diagnosis))),1)
- transform(as.data.frame(table(hosp$Diagnosis)),percentage_column=Freq/nrow(hosp)*100)

Result:

```
"Fracture"
```

Explanation:
First, let's join the two datasets (NEISS2014 and DiagnosisCodes) to display the diagnosis names. Since, the disposition code for hospitalization is 4, we can subset the dataset using the equal to operator. Then, we can group the types of diagnosis and calculate their frequency as well as the percentage rate to know the diagnosis corresponding to the highest number.

b) What diagnosis most often concluded with the individual leaving without being seen?

Code:

- Diagno <- left_join(NEISS2014, DiagnosisCodes, by = c("diag" = "Code"))

```
leave <- subset(Diagno , Diagno$disposition == 6)
tail(names(sort(table(leave$Diagnosis))),2)
tail(names(sort(table(leave$diag_other))),2)
```

Result:

```
"Laceration"            "Other/Not Stated"

"PAIN"                  ""
```

Explanation:
First, let's join the two datasets (NEISS2014 and DiagnosisCodes) to display the diagnosis names. Since, the disposition code for individual leaving without being seen is 6, we can subset the dataset using the equal to operator. Then, we can group the types of diagnosis and calculate their frequency know the diagnosis corresponding to the highest number.

The most often diagnosis comes out to be "Other/Not Stated". So, with further investigation, it can be seen that **"Laceration"** and **"Pain"** both are the most often diagnosis where the individual leaves without being seen.

c) Briefly discuss your findings and any caveats you'd mention when discussing this data

- People who had fractures, had to be hospitalized for the future treatment.
- People in general, who were suffering from pain or deep cuts(Laceration) left without much ado.
- The meaning for "hospitalization rate" seemed to be vague. It was not clear whether the disposition = 4 was the only case to be considered or not.

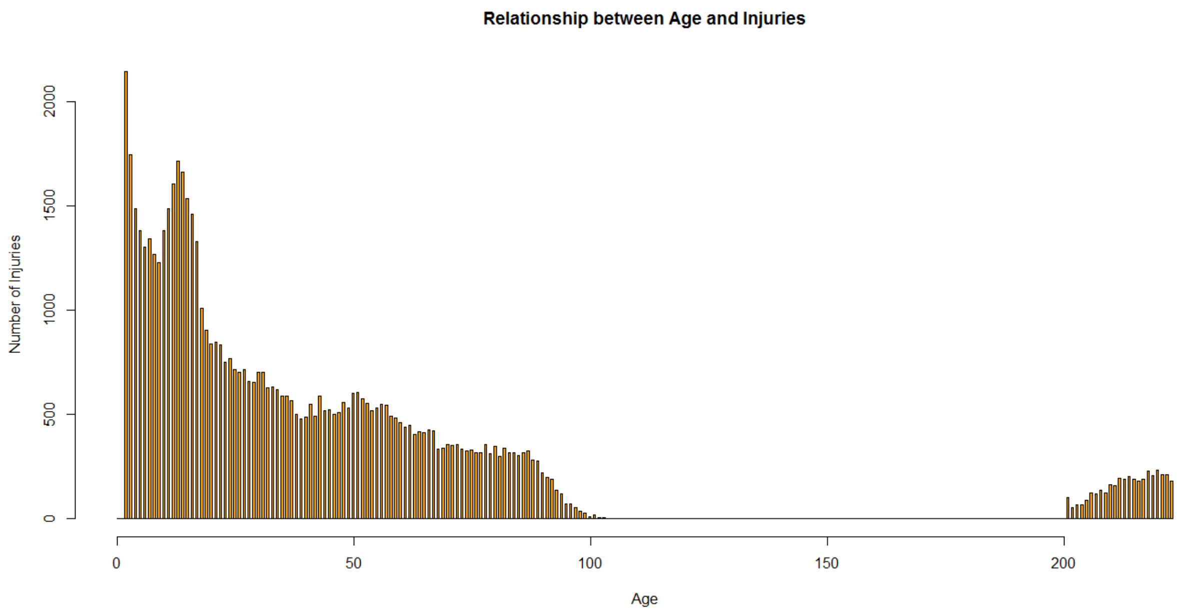
4) Visualize any existing relationship between age and reported injuries

Code:

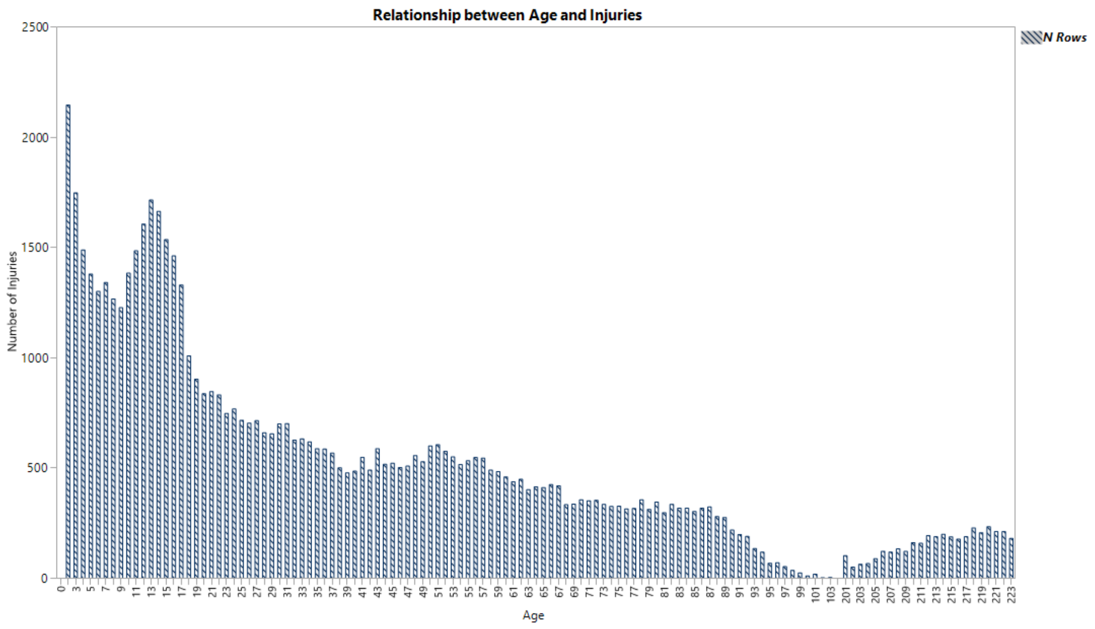
```
hist(NEISS2014$age, col="orange", main = "Relationship between Age and Injuries"
, breaks = 500, xlab = "Age", ylab = "Number of Injuries")
```

Explanation:
The below graphs are visualizations showing the number of injuries with respect to the age (Same using R and JMP). It can be seen that the maximum injuries recorded were at the age of 2. Then the number of injuries decreased and again peaked between the age of 10 – 17 years. From the age of 18, the number of injuries had a slow decrease, peaking a little at 30-31, 41, 43, 49- 53, 56-57. Also, it can be seen that comparatively there were less incidents below the age of 2.

The graph makes sense, as kids below 2 years are generally taken utmost care which implies they are prone to have less injuries. At 2 years, a possible explanation could be that, since kids start walking/running on their own and are generally admitted into day-cares, they are prone to experience a huge amount of incidents involving injuries. Later, we can say that teens are more prone to have accidents, as they now start using a lot of products and are less cautious at that age. As age increases, people become more cautious which decreases the risk of having injuries. Probably, at middle-age due to health problems we can see the little peaks in the graph.



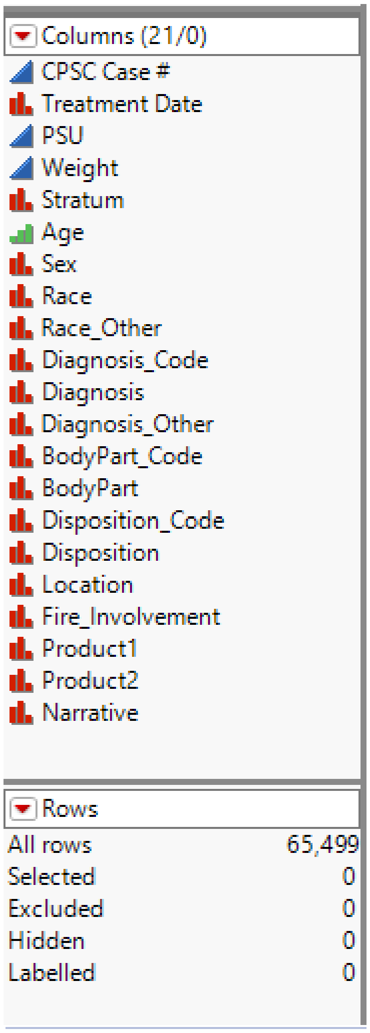
Using R



Using JMP

5) Investigate the data however you like and discuss any interesting insights you can find in the data

I have imported and joined all the four datasets in JMP.
Also, I have changed the “Data Type” and “Modelling Type” of few variables in the dataset for the below visualizations.
This is because, Ordinal/Nominal gives the frequency counts and Continuous give the average calculations.
Below is the current dataset variable list:



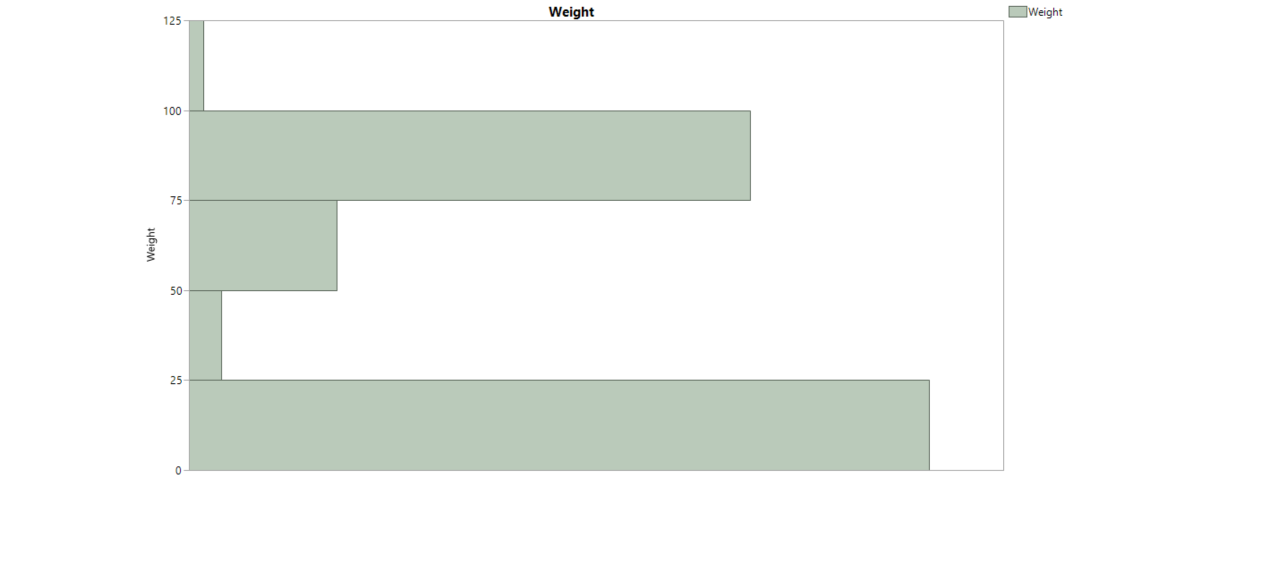
Columns (21/0)

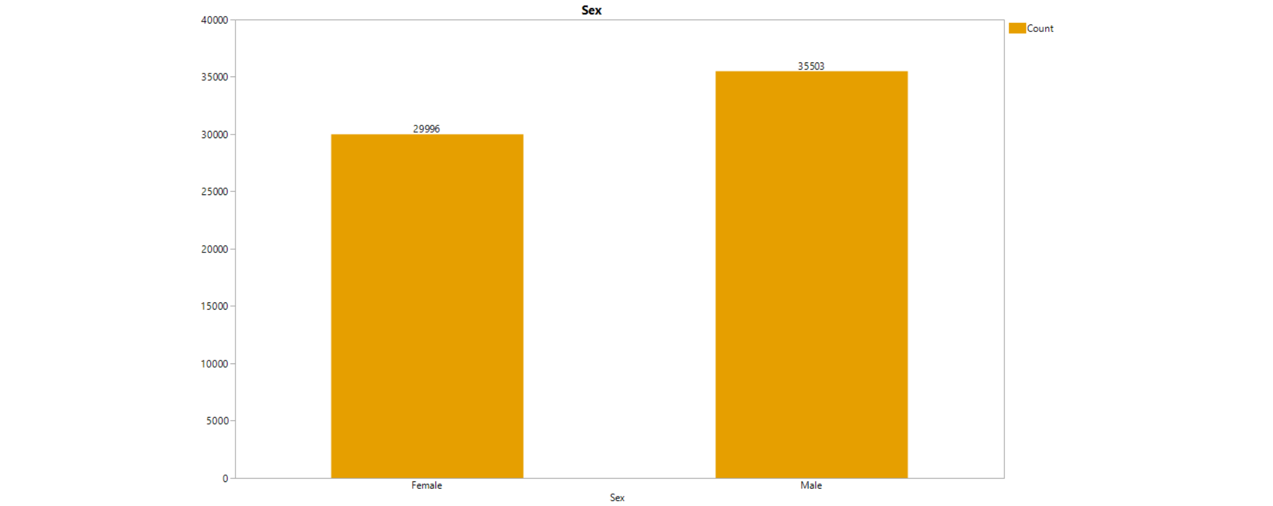
CPSC Case #
Treatment Date
PSU
Weight
Stratum
Age
Sex
Race
Race_Other
Diagnosis_Code
Diagnosis
Diagnosis_Other
BodyPart_Code
BodyPart
Disposition_Code
Disposition
Location
Fire_Involvement
Product1
Product2
Narrative

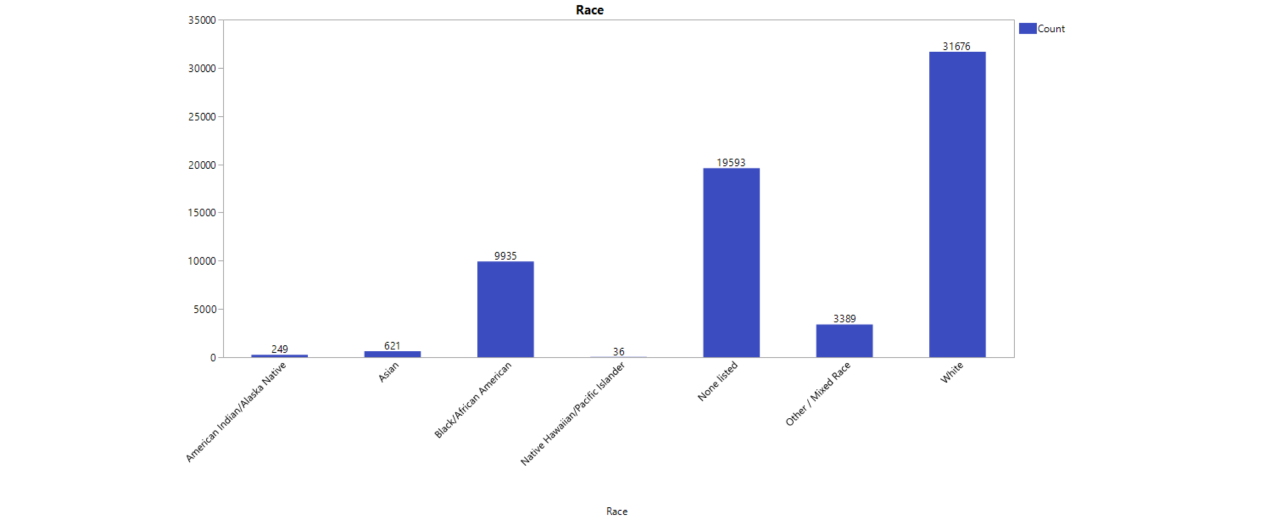
Rows

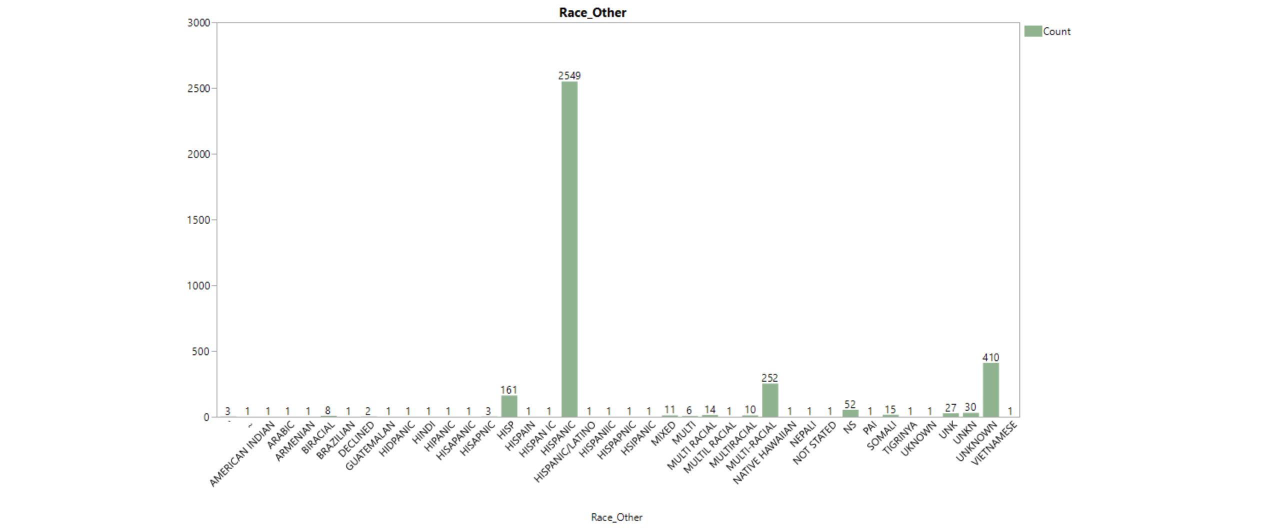
All rows	65,499
Selected	0
Excluded	0
Hidden	0
Labelled	0

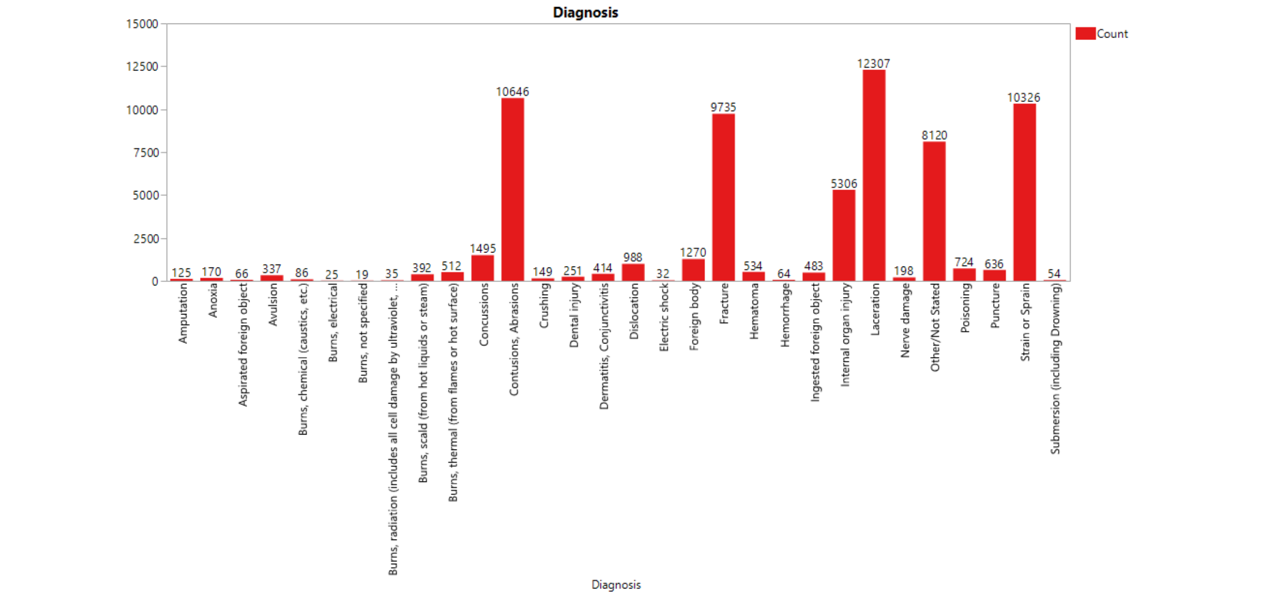
Investigating individual variables with respect to the number of Injuries:

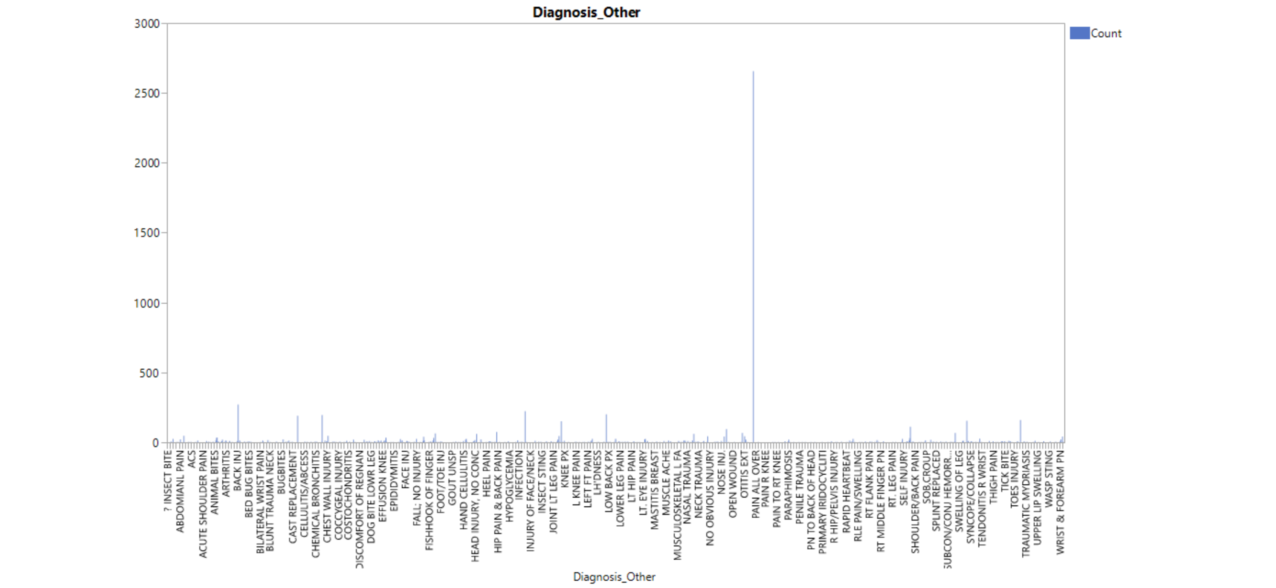


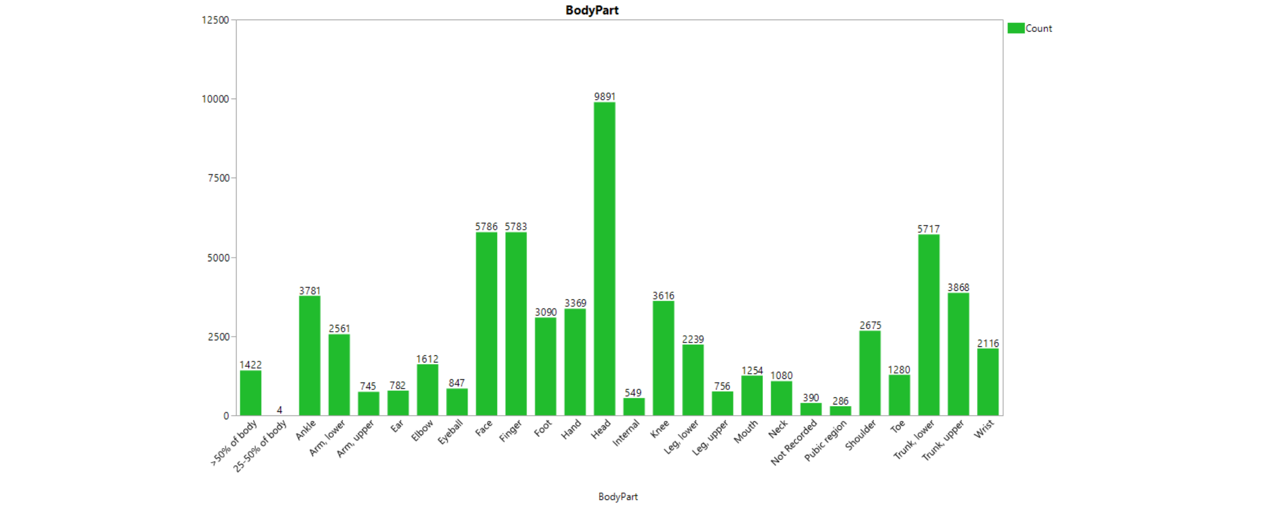


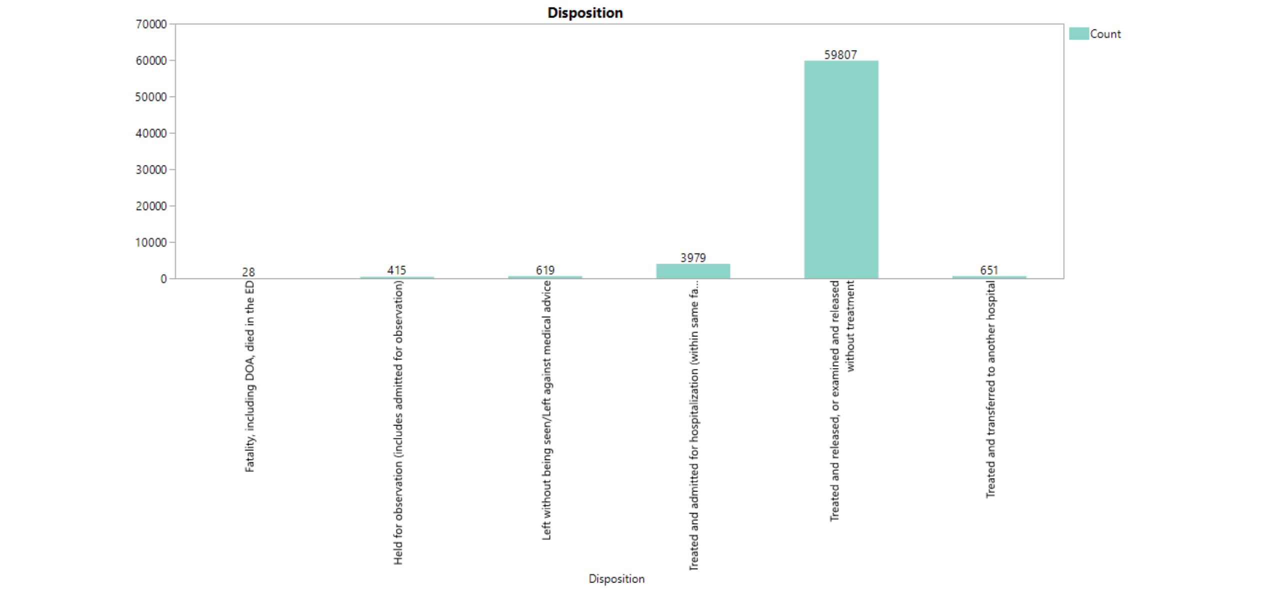


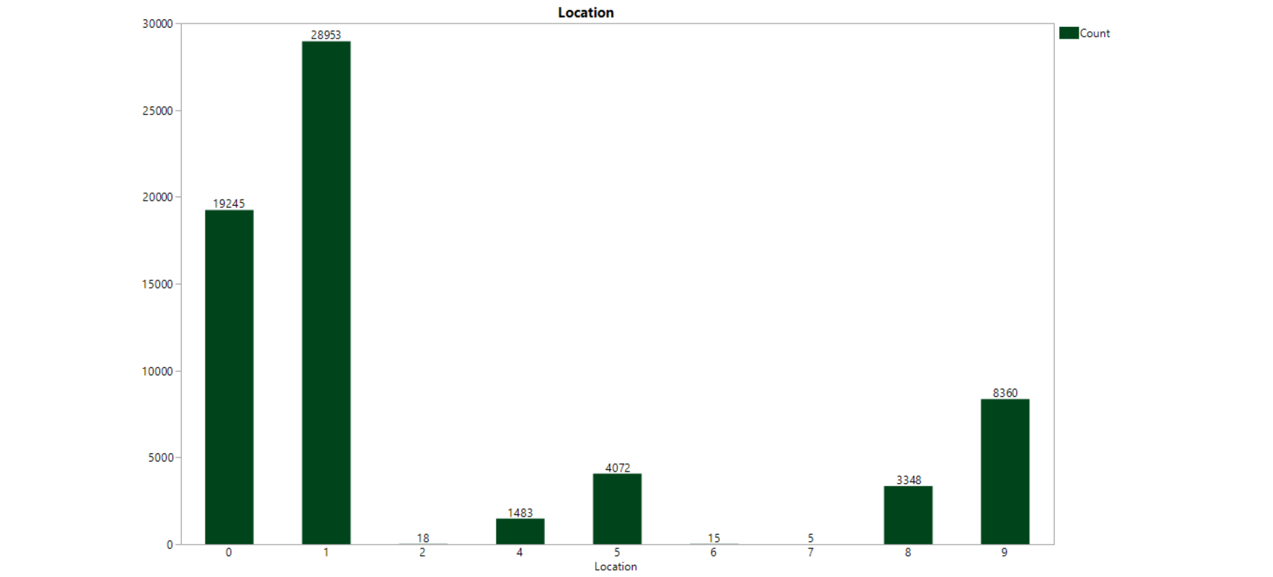

Race_Other
Count
3000
2500
2000
1500
1000
500
0
3
1
1
1
1
8
1
2
1
1
1
1
1
3
161
1
1
2549
1
1
1
1
11
6
14
1
10
252
1
1
1
52
1
15
1
1
27
30
410
1
-
~
AMERICAN INDIAN
ARABIC
ARMENIAN
BIRACIAL
BRAZILIAN
DECLINED
GUATEMALAN
HIDPANIC
HINDI
HIPANIC
HISAPANIC
HISAPNIC
HISP
HISPAIN
HISPAN IC
HISPANIC
HISPANIC/LATINO
HISPANIIC
HISPAPNIC
HSIPANIC
MIXED
MULTI
MULTI RACIAL
MULTIL RACIAL
MULTIRACIAL
MULTI-RACIAL
NATIVE HAWAIIAN
NEPALI
NOT STATED
NS
PAI
SOMALI
TIGRINYA
UKNOWN
UNK
UNKN
UNKNOWN
VIETNAMESE
Race_Other

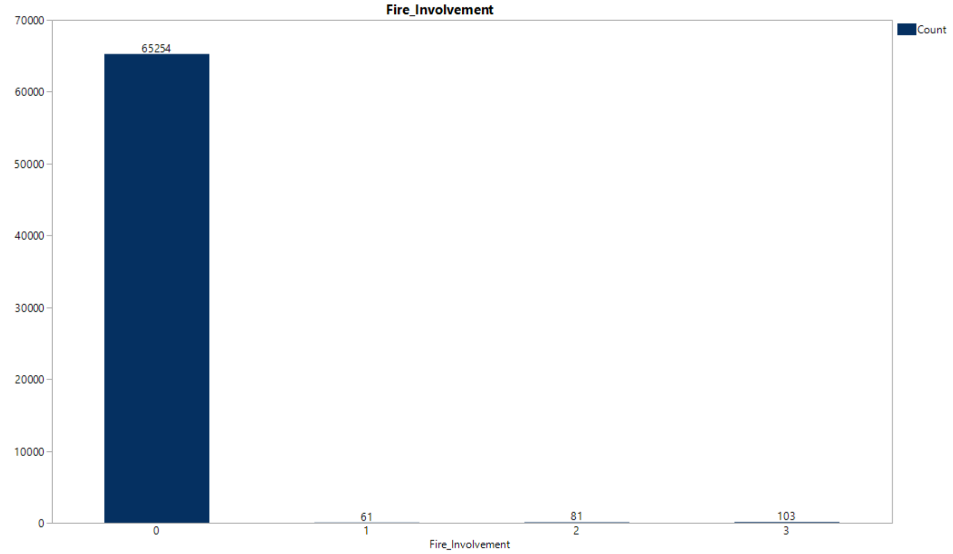

Diagnosis
Count
15000
12500
10000
7500
5000
2500
0
125
170
66
337
86
25
19
35
392
512
1495
10646
149
251
414
988
32
1270
9735
534
64
483
5306
12307
198
8120
724
636
10326
54
Amputation
Anoxia
Aspirated foreign object
Avulsion
Burns, chemical (caustics, etc.)
Burns, electrical
Burns, not specified
Burns, radiation (includes all cell damage by ultraviolet, ...
Burns, scald (from hot liquids or steam)
Burns, thermal (from flames or hot surface)
Concussions
Contusions, Abrasions
Crushing
Dental injury
Dermatitis, Conjunctivitis
Dislocation
Electric shock
Foreign body
Fracture
Hematoma
Hemorrhage
Ingested foreign object
Internal organ injury
Laceration
Nerve damage
Other/Not Stated
Poisoning
Puncture
Strain or Sprain
Submersion (including Drowning)
Diagnosis

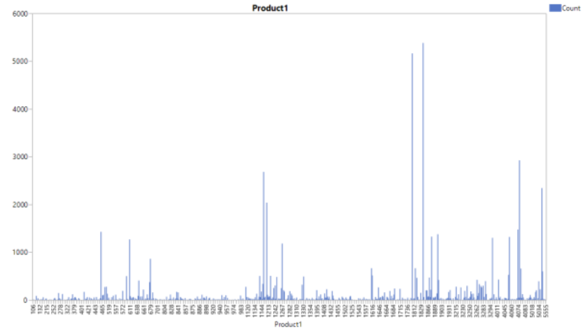

Diagnosis_Other
Count
3000
2500
2000
1500
1000
500
0
? INSECT BITE
ABDOMIANL PAIN
ACS
ACUTE SHOULDER PAIN
ANIMAL BITES
ARTHRITIS
BACK INJ
BED BUG BITES
BILATERAL WRIST PAIN
BLUNT TRAUMA NECK
BUGBITES
CAST REPLACEMENT
CELLULITIS/ABCESS
CHEMICAL BRONCHITIS
CHEST WALL INJURY
COCCYGEAL INJURY
COSTOCHONDRITIS
DISCOMFORT OF REGNAN
DOG BITE LOWR LEG
EFFUSION KNEE
EPIDIDYMITIS
FACE INJ
FALL; NO INJURY
FISHHOOK OF FINGER
FOOT/TOE INJ
GOUT UNSP
HAND CELLULITIS
HEAD INJURY, NO CONC
HEEL PAIN
HIP PAIN & BACK PAIN
HYPOGLYCEMIA
INFECTION
INJURY OF FACE/NECK
INSECT STING
JOINT LT LEG PAIN
KNEE PX
L KNEE PAIN
LEFT FT PAIN
LH'DNESS
LOW BACK PX
LOWER LEG PAIN
LT HIP PAIN
LT. EYE INJURY
MASTITIS BREAST
MUSCLE ACHE
MUSCULOSKELETAL L FA
NASAL TRAUMA
NECK TRAUMA
NO OBVIOUS INJURY
NOSE INJ.
OPEN WOUND
OTITIS EXT
PAIN ALL OVER
PAIN R KNEE
PAIN TO RT KNEE
PARAPHIMOSIS
PENILE TRAUMA
PN TO BACK OF HEAD
PRIMARY IRIDOCYCLITI
R HIP/PELVIS INJURY
RAPID HEARTBEAT
RLE PAIN/SWELLING
RT FLANK PAIN
RT MIDDLE FINGER PN
RT. LEG PAIN
SELF INJURY
SHOULDER/BACK PAIN
SOB,CROUP
SPLINT REPLACED
SUBCON/CONJ HEMORR...
SWELLING OF LEG
SYNCOPE/COLLAPSE
TENDONITIS R WRIST
THIGH PAIN
TICK BITE
TOES INJURY
TRAUMATIC MYDRIASIS
UPPER LIP SWELLING
WASP STING
WRIST & FOREARM PN
Diagnosis_Other

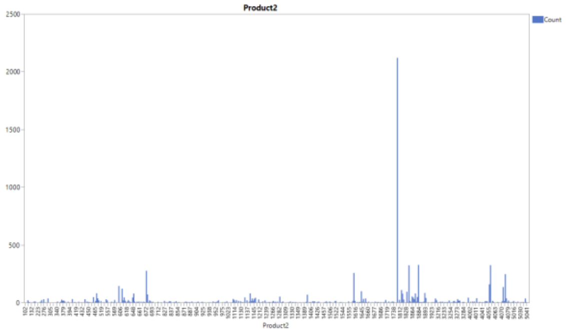

BodyPart
Count
12500
10000
7500
5000
2500
0
1422
4
3781
2561
745
782
1612
847
5786
5783
3090
3369
9891
549
3616
2239
756
1254
1080
390
286
2675
1280
5717
3868
2116
>50% of body
25-50% of body
Ankle
Arm, lower
Arm, upper
Ear
Elbow
Eyeball
Face
Finger
Foot
Hand
Head
Internal
Knee
Leg, lower
Leg, upper
Mouth
Neck
Not Recorded
Pubic region
Shoulder
Toe
Trunk, lower
Trunk, upper
Wrist
BodyPart

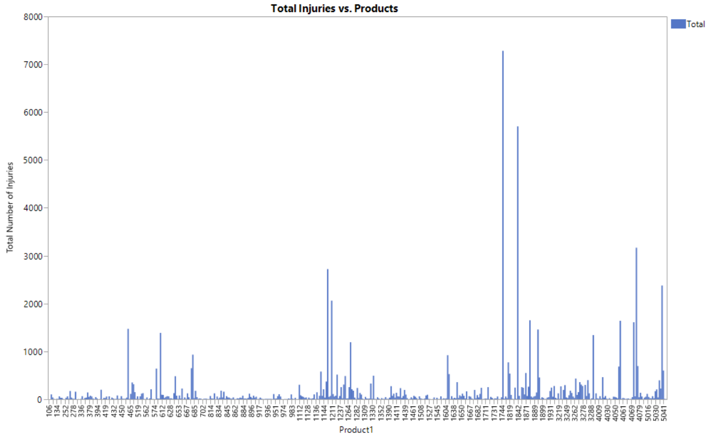

Disposition
Count
70000
60000
50000
40000
30000
20000
10000
0
28
415
619
3979
59807
651
Fatality, including DOA, died in the ED
Held for observation (includes admitted for observation)
Left without being seen/Left against medical advice
Treated and admitted for hospitalization (within same fa...
Treated and released, or examined and released without treatment
Treated and transferred to another hospital
Disposition


Location
Count
30000
25000
20000
15000
10000
5000
0
19245
28953
18
1483
4072
15
5
3348
8360
0
1
2
4
5
6
7
8
9
Location


Fire_Involvement
Count
70000
60000
50000
40000
30000
20000
10000
0
65254
61
81
103
0
1
2
3
Fire_Involvement


Product1
Count
6000
5000
4000
3000
2000
1000
0
106 132 215 252 278 322 379 401 421 443 465 519 557 572 611 638 661 679 701 804 828 841 857 875 886 898 920 940 957 974 983 1120 1134 1144 1213 1242 1267 1282 1310 1330 1354 1395 1408 1432 1455 1502 1525 1543 1557 1616 1646 1664 1684 1715 1736 1812 1829 1866 1889 1903 1931 3215 3230 3250 3262 3283 3294 4011 4045 4060 4074 4083 5018 5034 5555
Product1


Product2
Count
2500
2000
1500
1000
500
0
102 132 223 276 305 340 379 394 419 432 450 465 519 557 569 606 618 648 661 672 693 712 827 837 854 871 887 904 925 938 952 975 1023 1114 1130 1137 1145 1212 1239 1266 1282 1309 1330 1349 1389 1406 1426 1457 1506 1522 1544 1555 1616 1645 1660 1677 1686 1719 1738 1812 1828 1864 1884 1893 1923 3216 3233 3254 3273 3284 4002 4011 4041 4055 4063 4070 4079 5016 5030 5041
Product2


Total Injuries vs. Products
Total
Total Number of Injuries
8000
7000
6000
5000
4000
3000
2000
1000
0
106 134 252 278 336 379 394 419 432 450 465 519 562 574 612 628 653 667 685 702 814 834 845 862 884 896 917 936 951 974 983 1112 1128 1136 1144 1211 1237 1264 1282 1309 1330 1352 1390 1411 1439 1461 1508 1527 1545 1604 1638 1650 1667 1682 1711 1731 1744 1819 1842 1871 1889 1899 1931 3219 3249 3262 3278 3288 4009 4030 4050 4061 4069 4079 5016 5030 5041
Product1

Results:

- People weighing 0 – 25 Units have faced more injuries, followed by people weighing 75 – 100, 50 – 75, 25 – 50 and 100 – 125 Units. This matches with the age relationship.
- Males have a little more number of injuries as compared to Females. This is obvious, as their population is higher.
- White people have more injuries, followed by Black/African American people, followed by Hispanic people. This makes sense, as the data is from the hospitals in the US and its territories. Also, a major portion of injuries (19,593) have the race unlisted. So, this can be taken as a partial result.
- The top five list of diagnosis are:

  Laceration
  Contusions, Abrasion
  Strain or Sprain
  Fracture
  Internal Organ Injury
- Head is the most often injured part of the body. Face, finger and lower trunk parts of the body are also injured often, after head and almost in an equal proportion.
- Maximum proportion of the cases were treated and released or examined and released without treatment.
- It is surprising to see that most of the injuries happened in a farm/ranch followed by home; industrial place being the least.
- Majority of the incidents did not involve fire at all. The one's which involved fire had an approximately equal share of the fire department involvement (or not).
- Maximum injuries reported (7282) have been because of the Floors or Flooring materials, followed by Stairs or Steps (excluding pull-down and folding stairs).

  [**Note**: This was achieved by grouping by product 1 and product 2 and merging data to find out the total number of injuries caused by the product.

  Please find the attached dataset.]

Future Work:

Relationship between multiple variables and the number of injuries can be found out and visualized for more interesting facts.