Final Project

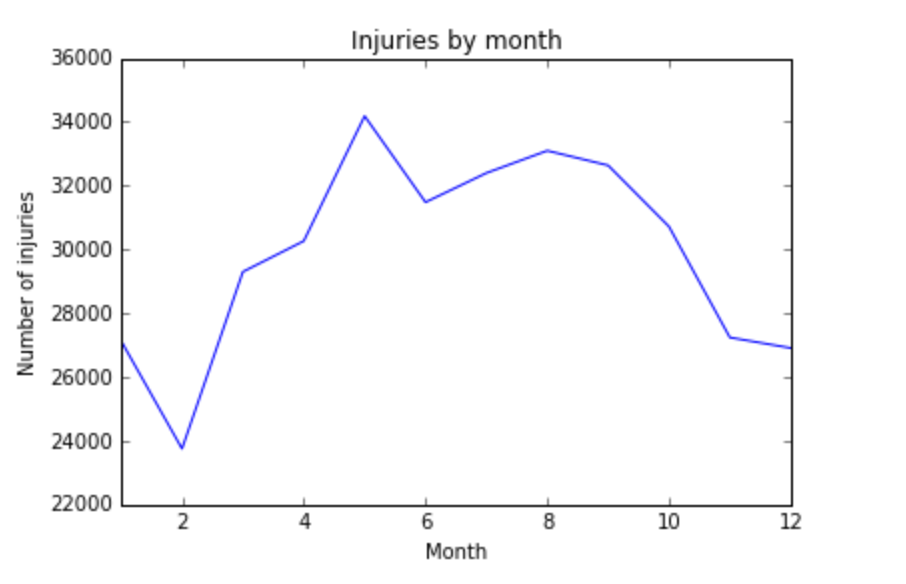
Peter Zhou, Omar Al Taher, Jen Jen Chen

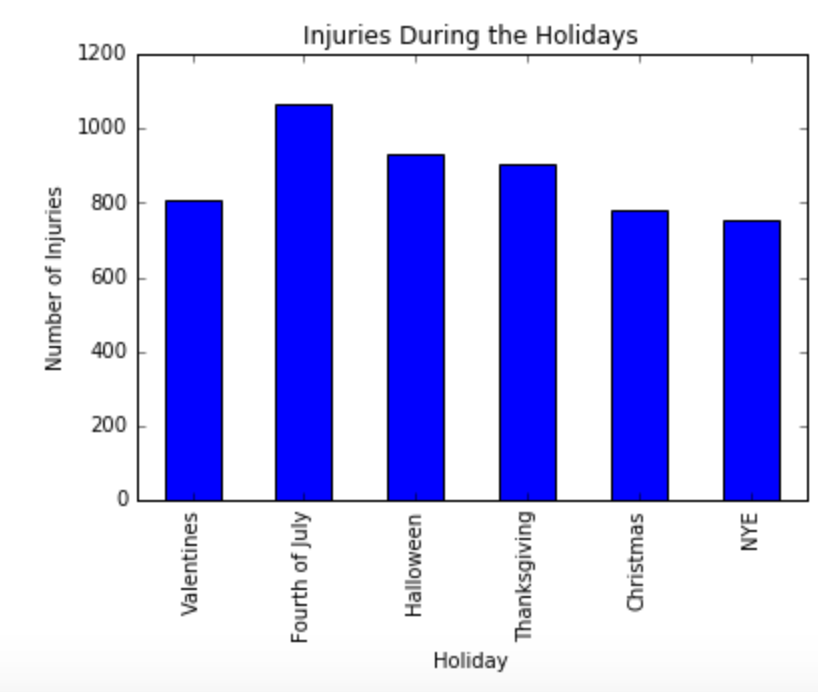
Our group chose the 2015 dataset from the National Electronic Injury Surveillance System (NEISS), which contained patient information for hospital emergency room visits in the United States. We initially began this journey with the following questions:

1. Which emotional state was the most and the least associated with injury?
2. Which household item pose the largest threat to its inhabitants?
3. Injuries that involve animals are most common and least common to which animal?
4. Which parts of the body are the most vulnerable to injury?
5. What types of locale are associated with the most injuries?
6. Of those injuries inflicted or received while under the influence, what proportion are related to alcohol versus drugs?
7. What are the peak times that injuries occur? Is there a seasonal pattern or monthly pattern?
8. Which sex is more prone to injuries?
9. What type of injury would be considered as “junk in the trunk” as referred to in the notes?

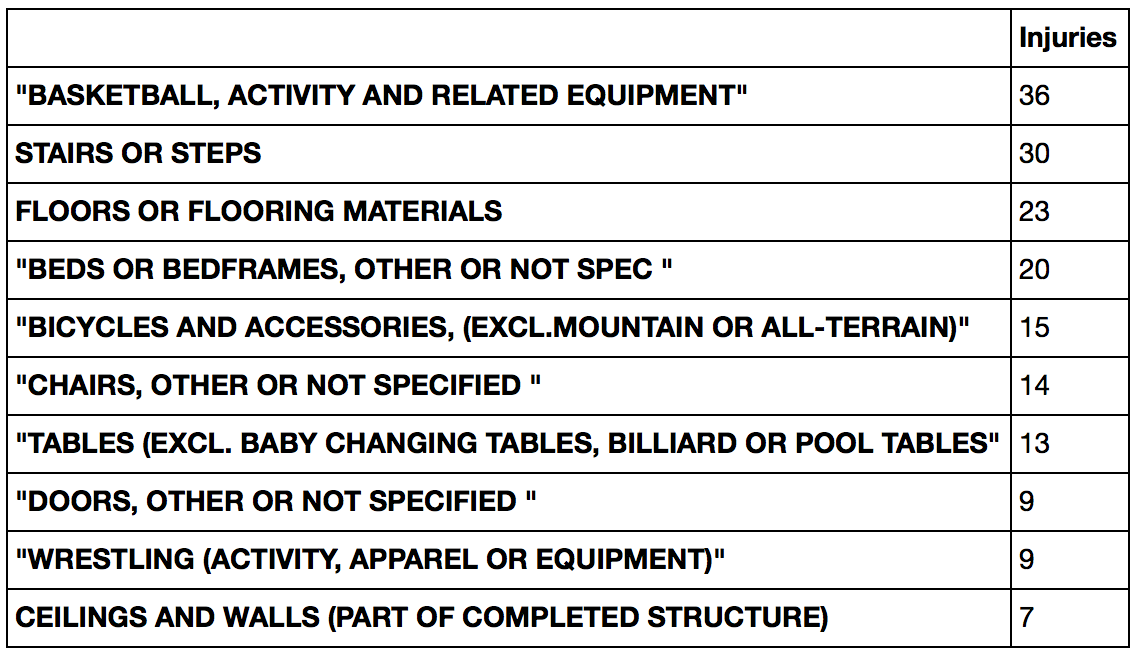
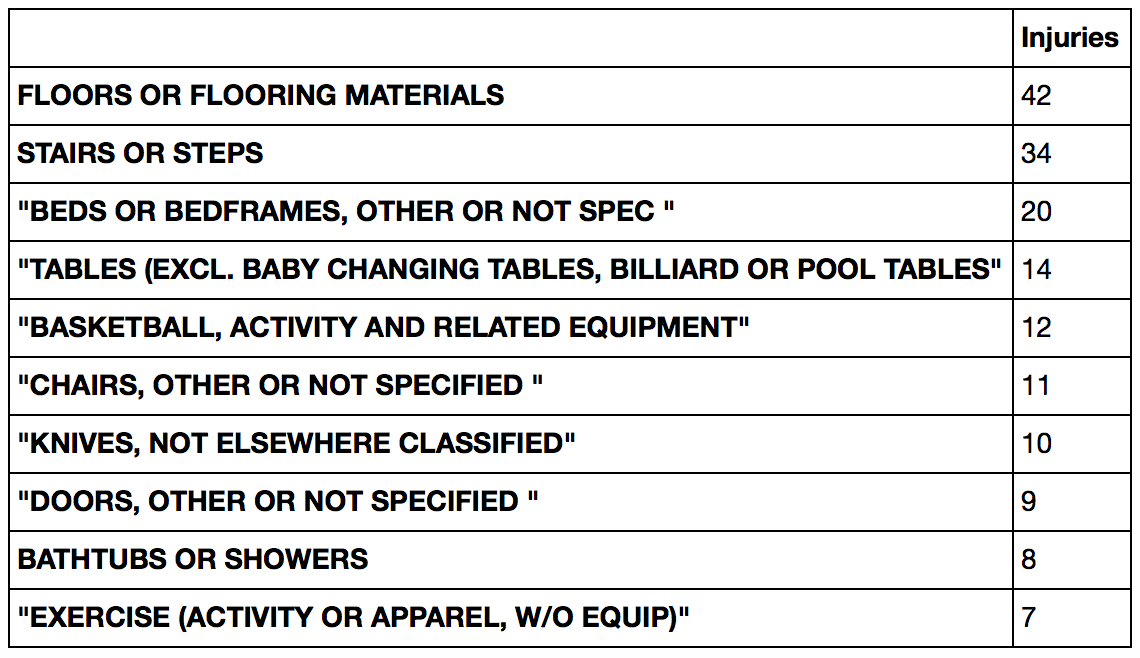
These initial questions were continually evolving as we began checking our data and testing outcomes from our analysis. Our goal was to provide the audience with valuable insights that would minimize their risk of injury, and we chose to modify our questions if we discovered a lack of significance in the results.

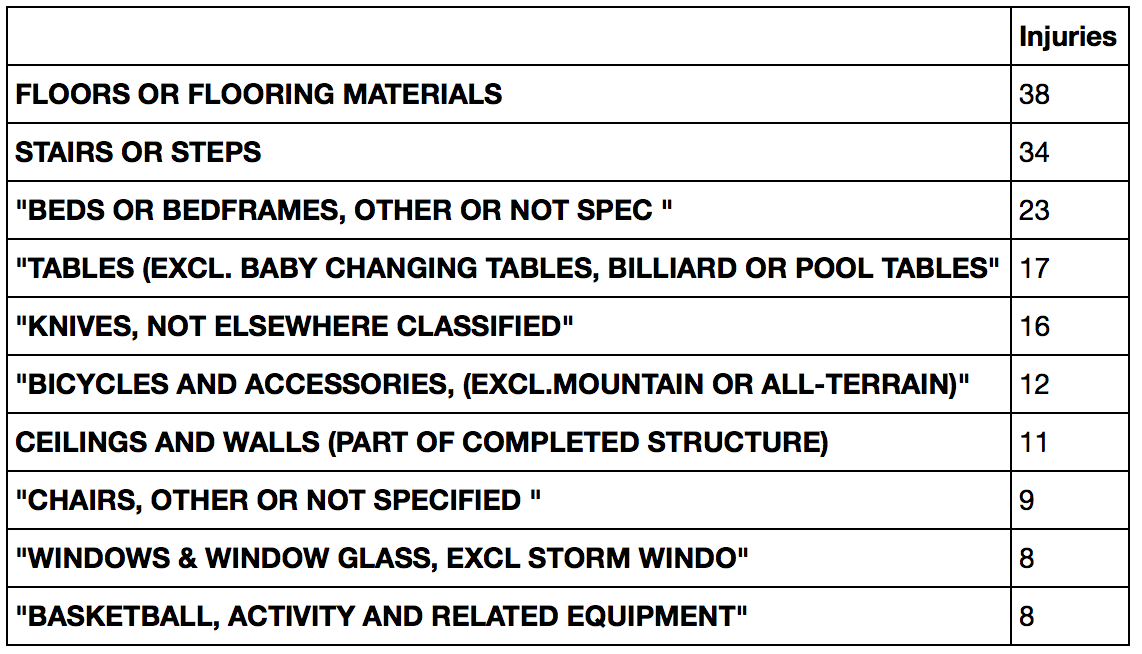
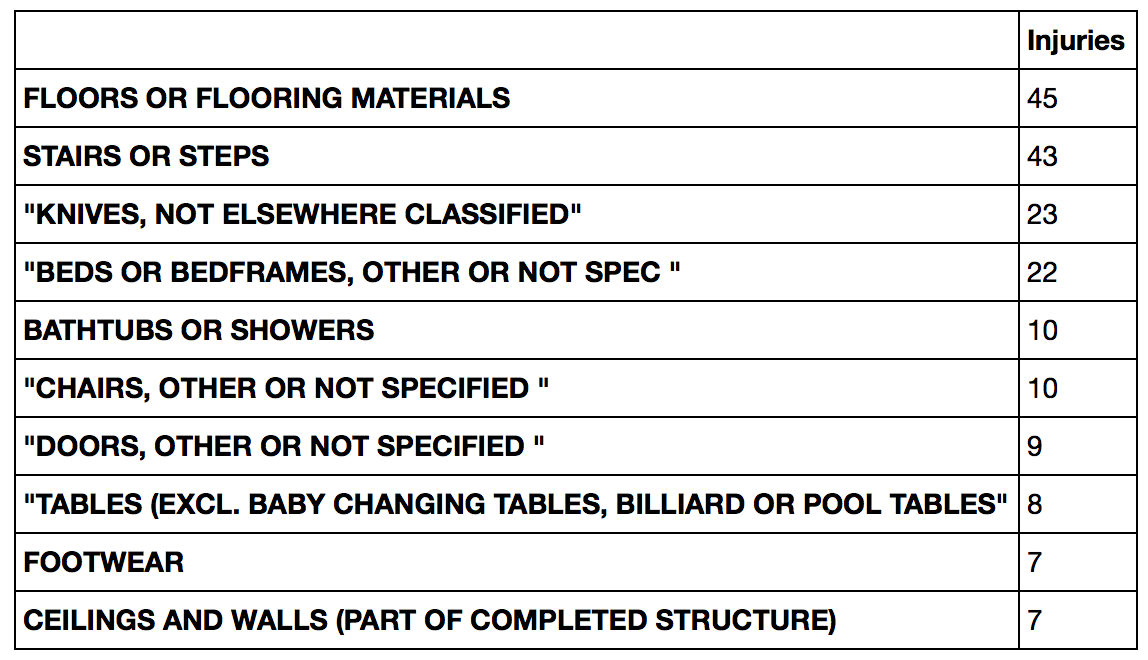
**Time-related injuries**

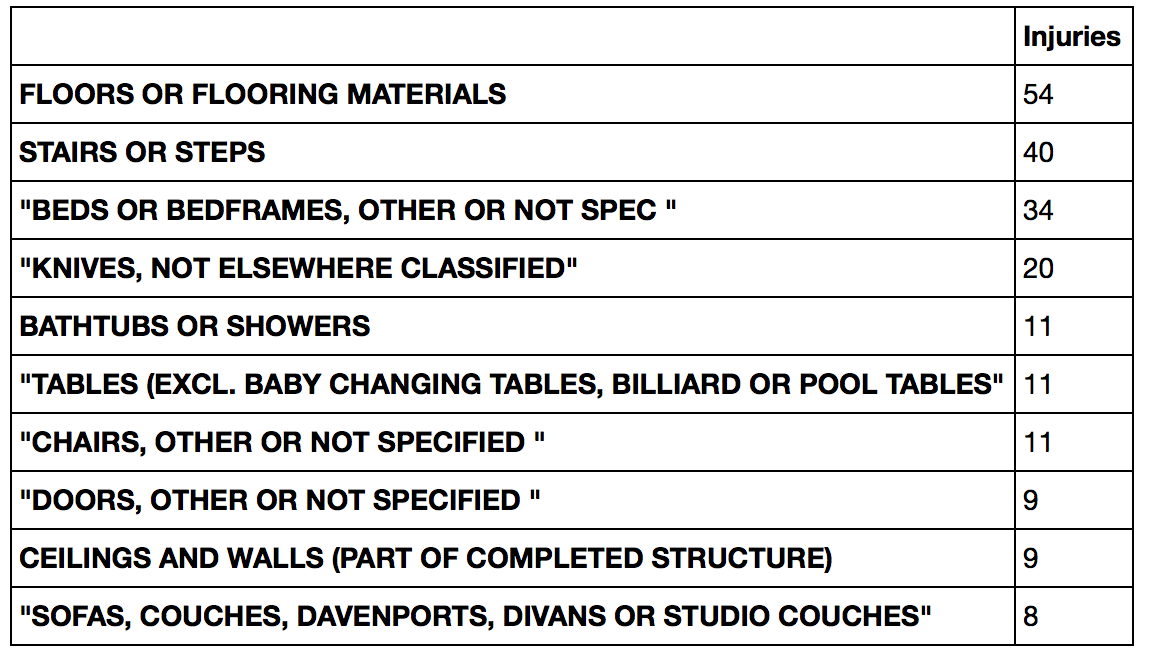
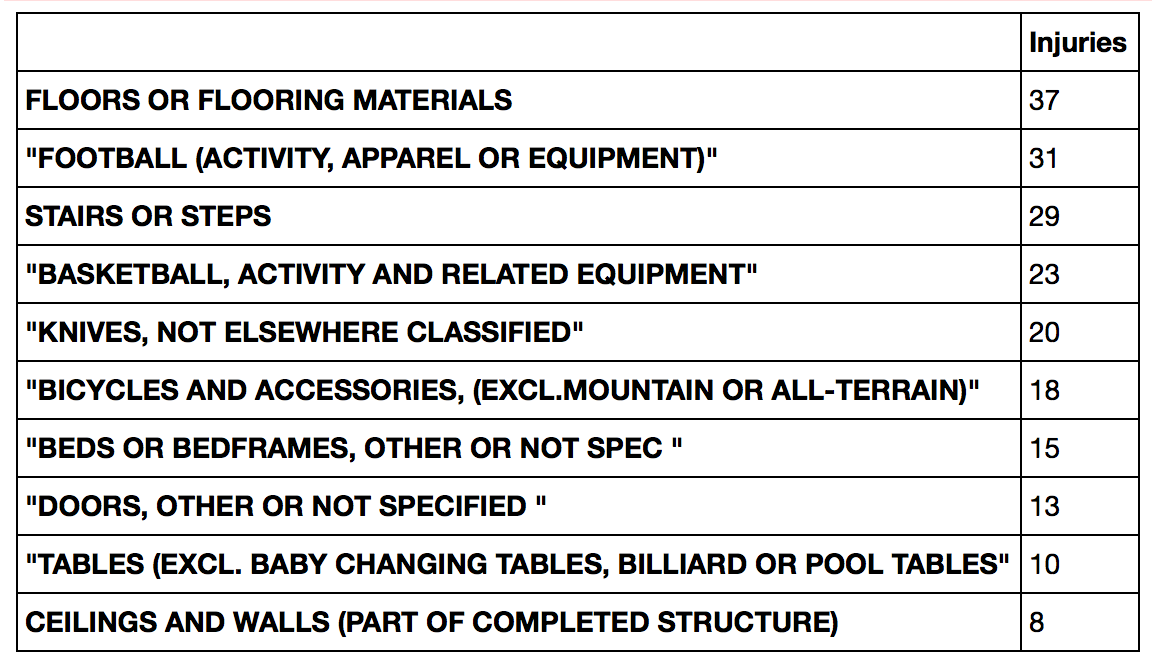
The process of looking for time-related injuries was a valuable lesson in panda’s DateTime formats and its manipulation. The most difficulty we had was being able to bin the dates into months and still be able to track the number of injuries for that month. Although it was relatively straightforward to compile each month’s injury records, it was an obstacle to be able to even count the number of occurrences, because using such functions as .count() (.value\_counts() was not able to be used) would output all the other columns (‘age’, ‘weight’, ‘sex\_descri’, etc) in as total numbers for each column instead. To solve this problem, we set the index to be the dates instead of having the dates in a column. By doing so, we were able easily use the groupby method to refer to index.month and count the number of occurrences.

The “Injuries by Month” graph depicts a spike in May and in September, with the lowest points from November to February. This pattern led us to consider adding on analysis comparing different holidays of the year. In order to create a graph of the major holidays - Valentine’s Day, July 4th, Halloween, Thanksgiving, Christmas, New Years Eve - we first had to set up each holiday’s analysis and the combine them into a single bar chart. Referring to each holiday was straightforward as we again used the index to refer to all injuries that occurred on specific dates. After which we took a count of each holiday. We were able to create a separate data frame using these injury counts and the holidays they occurred on.

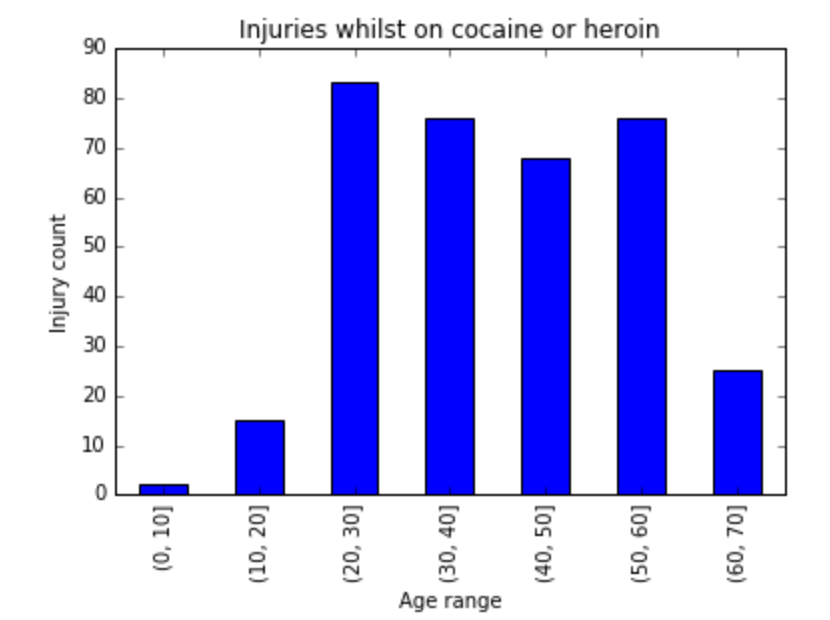
For each holiday, we wanted to take a closer look at the consumer products which were most commonly involved in injury and compare similarities and differences across the holidays. This sections required the most trial and error, in terms of trying out different grouping and ranking methods. It was challenging to try grouping it by age or by sex, and then creating graphs and plots from the grouped data. However, we found that graphing even just the most common consumer products by sex and by holiday was a poor depiction. Instead, by grouping it first by sex (‘sex\_descrip’) and then by product (‘product1\_descrip’), we were able to take the ten most common injuries using a value\_count method, and each of these counts were placed into a separate data frame to create a table which was easier to read than a graph The top ten male versus female consumer products were obtained for each holiday, and while “Floors or flooring materials” and “Stairs” were the most common objects, there were other several other observations that notably stood out (female tables on the right):

Valentine’s Day - females with knife-related injuries while males had basketball and wrestling injuries.

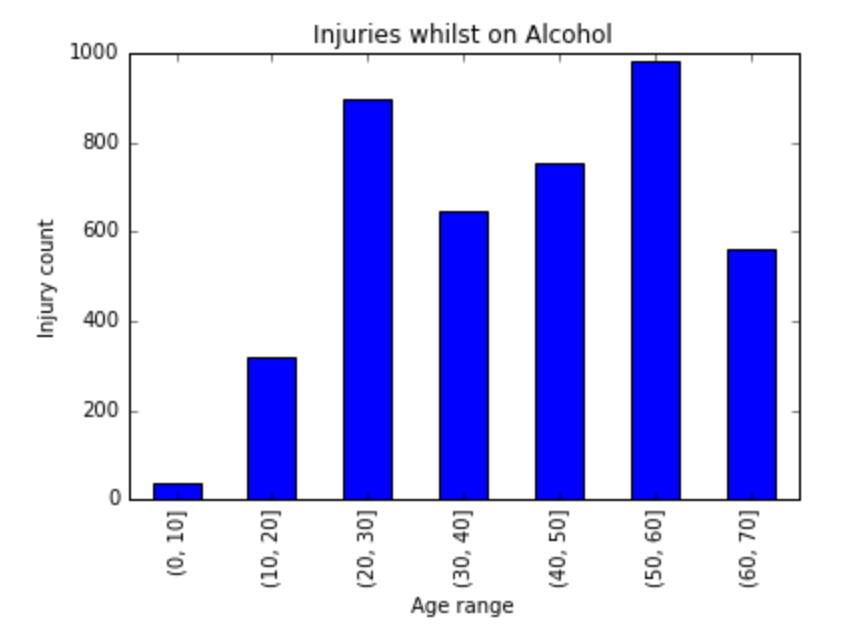
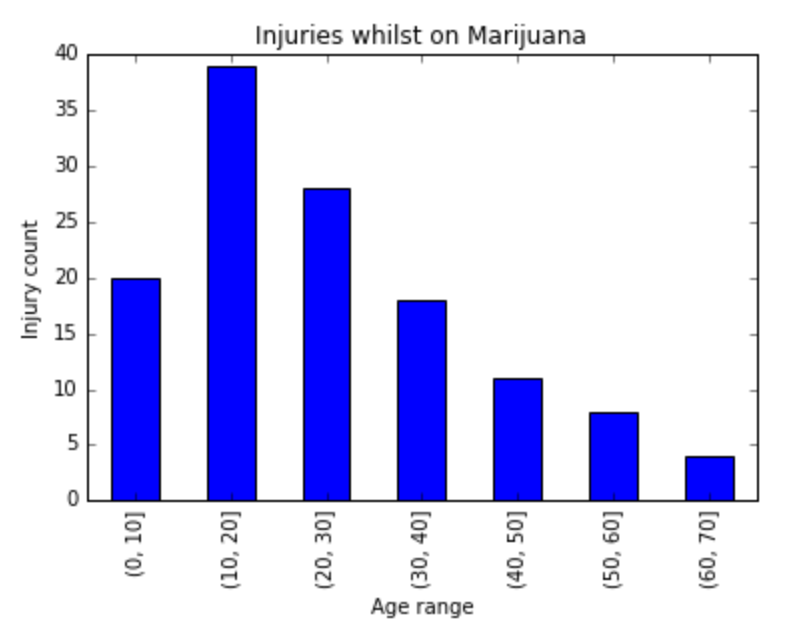
Christmas - females with knife-related injuries and males with bicycle and basketball-related injuries, although males do have some knife-related injuries. Perhaps they are helping in the kitchen more?

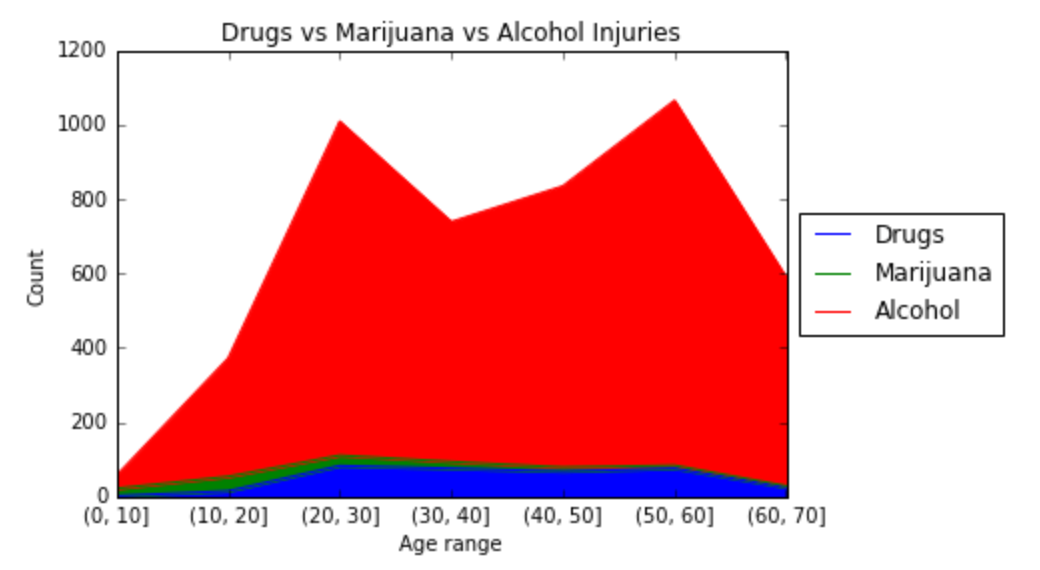
Thanksgiving - males with high football-related injur

Other observations included a substantial number of males with basketball-related injuries on New Year’s Eve, significantly more males with firework injuries on July Fourth than females,

**“Doped-up” Injuries**

The next section was looking in the miscellaneous section (‘Notes’) to compare injuries related to drugs, alcohol, or marijuana. Using the ‘Notes’ column, we were able to filter out the injury cases that had specific strings. For the drug evaluation, we used the str.contains method to look for “COCAINE|HEROIN’ and then stratify by age to obtain value counts. However, we found that illustrating it in a graph resulted in excessive number of bars, so we opted to bin the ages into groups of ten years and sum the value counts per bin to make a cleaner look. For each section, we used the minimum and maximum age indices as the start and stop point for the np.arange() function, then summed up the injury count per bin. This was done for drugs, alcohol, and marijuana.

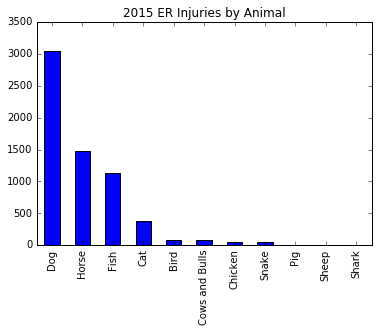
Filtering out the rows for the marijuana evaluation was the most time-consuming, because the words the the ER staff used commonly included “high” in the Notes column. This required checking each reference and adding a filter that would remove future rows with references to non-marijuana related products, such as “high heels” or “high school.” 

Finally, these three data frames were combined via concatenation to create a graph that would show any differences between the prevalence of a particular substance and what ages would be the most popular. We elected to use an area graph because it illustrated that alcohol was far by the most common substance, with peaks in the 20-30 and 50-60 years ranges, also that marijuana stemmed mostly from the younger (<30 yo) patients.

**Animal-related injuries**

One of the questions we were interested in answering is which animals cause people to go to the emergency room the most. Not surprisingly, people are injured by dogs the most. The assumption is that the animal humans are most exposed to are dogs. Surprisingly, the animal that causes the second most injuries are horses. The injuries caused by fish are mostly related to fishing. The safest safest household animal is a cat with only 372 incidents of cat related injuries in the year 2015 in our dataset of 359,129 ER visits.

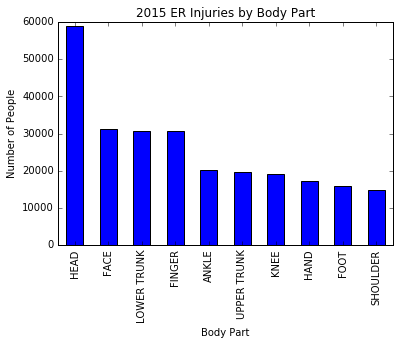
|  |  |
| --- | --- |
| Animal | Count |
| Dog | 3049 |
| Horse | 1475 |
| Fish | 1125 |
| Cat | 372 |
| Bird | 81 |
| Cows and Bulls | 77 |
| Chicken | 50 |
| Snake | 47 |
| Pig | 6 |
| Sheep | 2 |
| Shark | 1 |



Obtaining this data did have its limitations. We had to parse the data from the notes field by using keywords. This limits our findings to the combinations of words we could think of relating to each animal. We also had to eliminate occurrences that are unrelated to this question. For example, when looking up injuries caused by chickens, we had to remove anything related to cooking.

The most interesting animal related injury is a patient who “71YOM FELL OFF A HORSE& A COW STEPPED ON HIM”.

**Top 10 Body Parts Injuries**

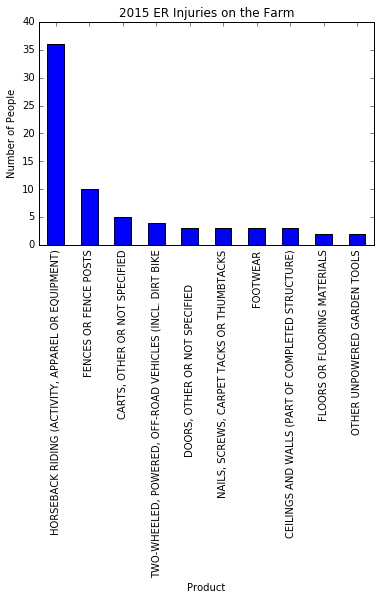
These are the top ten body parts that are injured and results in emergency room visits. The most common body part that is injured is the head followed by the face, lower back, and finger. The interesting take away is the disparity in age between the different body parts that are injured. The oldest category is the lower trunk. This consists mostly of lower back pain and lumbar issues. The data validates the general assumption that the older a person is, the more likely they are to have injuries sustained to their lower back. The face is the youngest category in the top ten. Most of the injuries sustained in this category involve physical activity or running around and falling on their face.

**Farm Injuries**

|  |  |  |
| --- | --- | --- |
| Farm Activity | Median Age | Count |
| HORSEBACK RIDING (ACTIVITY, APPAREL OR EQUIPMENT) | 24.5 | 36 |
| FENCES OR FENCE POSTS | 26 | 10 |
| CARTS, OTHER OR NOT SPECIFIED | 10 | 5 |
| TWO-WHEELED, POWERED, OFF-ROAD VEHICLES (INCL. DIRT BIKE | 21.5 | 4 |
| DOORS, OTHER OR NOT SPECIFIED | 15 | 3 |
| NAILS, SCREWS, CARPET TACKS OR THUMBTACKS | 32 | 3 |
| FOOTWEAR | 14 | 3 |
| CEILINGS AND WALLS (PART OF COMPLETED STRUCTURE) | 23 | 3 |
| SPORTS AND RECREATIONAL ACTIVITY, N.E.C. | 20 | 2 |
| LADDERS, OTHER OR NOT SPECIFIED | 11 | 2 |

One of the more interesting locations where injuries occurred was on a farm. The second most injuries caused by an animal came from a horse as shown from our earlier question of which animal causes the most injuries. Horseback riding is the leading cause of emergency room visits from a farm. Another dangerous place to avoid on a farm is the fence. Numerous injuries occurred with people falling onto and off of fences, getting caught in fence doors, or touching electric fences.

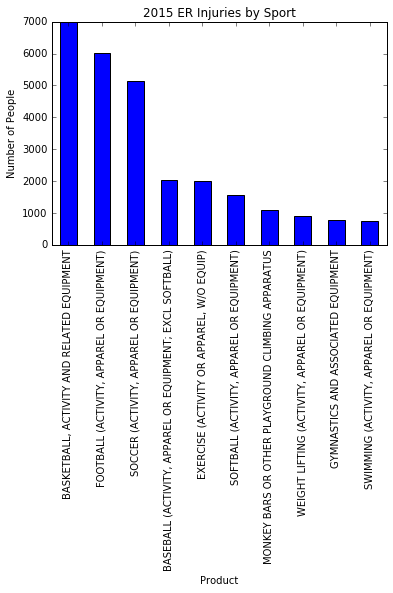
|  |  |  |
| --- | --- | --- |
| Body Part | Median Age | Count |
| HEAD | 36 | 58912 |
| FACE | 16 | 31104 |
| LOWER TRUNK | 48 | 30704 |
| FINGER | 26 | 30594 |
| ANKLE | 21 | 20148 |
| UPPER TRUNK | 46 | 19578 |
| KNEE | 27 | 19029 |
| HAND | 27 | 17140 |
| FOOT | 25 | 15961 |
| SHOULDER | 33 | 14874 |



**Sport Injuries**

|  |  |  |
| --- | --- | --- |
| Sport | Median Age | Count |
| BASKETBALL, ACTIVITY AND RELATED EQUIPMENT | 16 | 6987 |
| FOOTBALL (ACTIVITY, APPAREL OR EQUIPMENT) | 14 | 6034 |
| SOCCER (ACTIVITY, APPAREL OR EQUIPMENT) | 14 | 5127 |
| BASEBALL (ACTIVITY, APPAREL OR EQUIPMENT; EXCL SOFTBALL) | 14 | 2022 |
| EXERCISE (ACTIVITY OR APPAREL, W/O EQUIP) | 34 | 1999 |
| SOFTBALL (ACTIVITY, APPAREL OR EQUIPMENT) | 17 | 1558 |
| MONKEY BARS OR OTHER PLAYGROUND CLIMBING APPARATUS | 6 | 1083 |
| WEIGHT LIFTING (ACTIVITY, APPAREL OR EQUIPMENT) | 29 | 894 |
| GYMNASTICS AND ASSOCIATED EQUIPMENT | 11 | 767 |
| SWIMMING (ACTIVITY, APPAREL OR EQUIPMENT) | 15 | 748 |

The sport that caused the most injuries is basketball, followed by football and soccer. The average age of the top sports injuries are all teenagers. We also looked at the top ten injured body parts that are obtained from sports. Ankle injuries in basketball are the most prevalent followed by head injuries in football.



**Emotional Injuries**

As people's judgment can be impaired by their emotional state, the data was analyzed to find which emotions are most mentioned in the 'Notes' field.

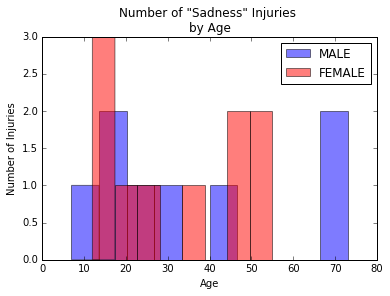
**Happiness**

Not surprisingly, no records were found to include the words "happy", "happiness", or "glad". One record did include "not happy", but that was not included in the count for obvious reasons.

|  |  |  |  |
| --- | --- | --- | --- |
| Sport | Body Part | Median Age | Count |
| BASKETBALL | ANKLE | 17 | 1482 |
| FOOTBALL | HEAD | 13 | 1046 |
| BASKETBALL | FINGER | 14 | 985 |
| FOOTBALL | FINGER | 14 | 789 |
| SOCCER | HEAD | 14 | 782 |
| BASKETBALL | KNEE | 17 | 775 |
| SOCCER | ANKLE | 15 | 744 |
| SOCCER | KNEE | 16 | 631 |
| FOOTBALL | KNEE | 15 | 630 |
| BASKETBALL | HEAD | 14 | 611 |

**Sadness**

Sadness returned 18 results, based on searching for "sad", "sadness", "depressed" or "unhappy". Records had to be filtered for the use of "depressed" to describe some injuries like "depressed skull", and records that indicated that the the patient was sad to be in the ER, which means that sadness was not the cause of the injury.

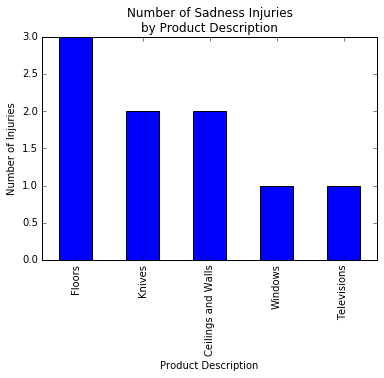


Injuries related to sadness affected females a bit more than males, as evidenced by the percentages of the injuries affecting each sex.

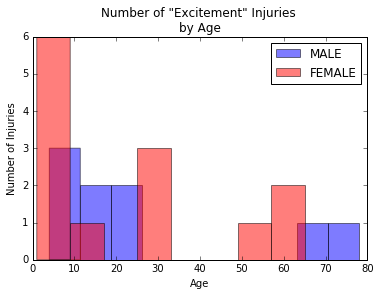
The mean age for these injuries was similar for males and females, being in the early thirties for both. When analyzed by age, a peak in injuries is seen in the teenage years for both sexes. This is the only observable pattern, which could be due to the relatively low number of data points or may be an indicator that sadness related injuries occur at a random rate across all ages beyond the teenage years.

|  | **Mean age** | **% of Injuries** |
| --- | --- | --- |
| **FEMALE** | 32.0 | 55.6 |
| **MALE** | 34.8 | 44.4 |

The top three products involved in the injuries related to sadness are floors, knives, and walls. By looking at the data, in the case of floors and walls, this is due to many people being found laying on the floor after they have been injured, or injuring themselves by bumping into walls. This might indicate that the primary cause is lack of focus or distraction due to being depressed or sad. In this dataset the knife injuries were evenly split between one due to distraction, and one due to self-harm.

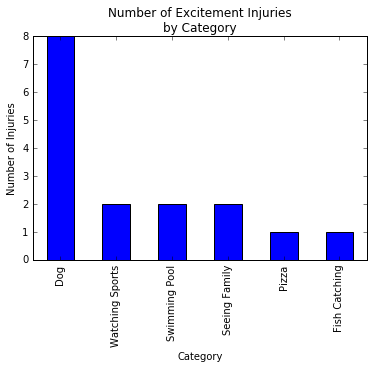


**Excitement**

Excitement returned 20 results, based on searching for "excited" or "excitement".

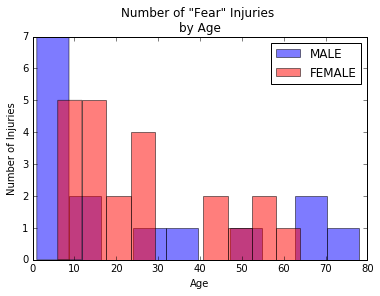
Injuries related to excitement affected females 50% more than males, as evidenced by the percentages of the injuries affecting each sex. The mean age for these injuries was similar for males and females, being in the mid twenties for both. When analyzed by age, a peak in injuries is seen in children through ten years old and younger for both sexes. The prevalence of these injuries falls dramatically with age, with 75% of injuries reported in people 40 years old or younger. This may be an indication that people older that 40 are less excitable, have less exciting lives, or have learned to curb their enthusiasm.

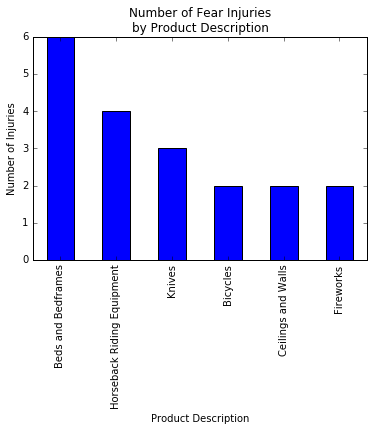
|  | **Mean age** | **% of Injuries** |
| --- | --- | --- |
| **FEMALE** | 22.5 | 59.1 |
| **MALE** | 26.2 | 40.9 |



The data was analyzed to find out what was causing the patient to get excited at the time of their injury, and sorted to find the most common causes.  
Dogs are the mentioned in more than 30% of the cases. All of these incidents involved a dog getting excited and biting a person (e.g. 32YOF LAYING DOWN WITH HER DOG AND THE DOORBELL RANG THE DOG BECAME EXCITED AND JUMPED ONTO PTS FACE).  
Watching sports, swimming pools, and seeing family were each reported twice, or in 10% of the cases.

**Fear**

Fear returned 37 results, based on searching for "fear", "afraid", "scared" or "frightened".

Injuries related to fear affected females 50% more than males, as evidenced by the percentages of the injuries affecting each sex. The mean age for these injuries was similar for males and females, being in the mid twenties for both. When analyzed by age, a peak in injuries is seen in children through their teenage years for both sexes. Lower peaks show up in later decades, but are scattered and seem random. This could be due to the relatively low number of data points or may be an indicator that fear-related injuries occur at a random rate across all ages beyond the teenage years.

|  | **Mean age** | **% of Injuries** |
| --- | --- | --- |
| **FEMALE** | 25.6 | 59.5 |
| **MALE** | 24.7 | 40.5 |

The top three products involved in the injuries related to fear are beds/bedframes, horseback riding equipment, and a tie in third for bicycles, ceilings/walls, knives, and fireworks.

The injuries related to beds are related to people getting startled in their beds then getting injured, or people falling out of them (e.g. "64 YO F C/O CHEST PAIN TODAY SAW A RAT IN HER ROOM GOT SCARED JUMPED OUT FROM HER BED HITTING CHEST ON DRESSING TABLE DX MSK PAIN" or "55 YR OLD FEMALE SCARED BY CAT AND FELL OUT OF BED CONTUSING KNEE”).

The injuries related to horseback riding equipment are all due to horses getting scared while someone was riding them, or after someone falls off of them (e.g. "27YOM AT THE PARK HORSEBACK RIDING WHEN A RATTLESNAKE SCARED HORSE,PT FELL FROM HORSE ON R HIP DX ACUTE R HIP CONTU,R THIGH CONTU,LUBMAR SPRAI"). It seems like the there is nothing to fear while horse riding, except the horse getting scared.

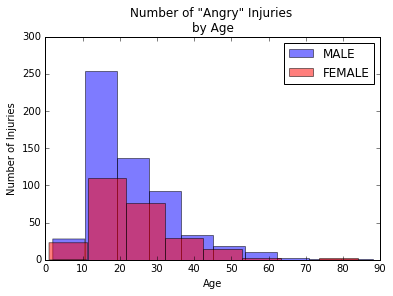
Thirdly, injuries related to knives are due to people getting startled while they were cutting something with a knife. Injuries related to fireworks (perhaps as some form of cosmic karma) were both due to the dogs being afraid of fireworks then biting a person (e.g. "19YOF ARM ABRASION WHEN BITEN BY DOG WHEN SCARED OF FIREWORKS. TYPE M FIREWORKS. DX ARMABRASION").

|  | **Mean age** | **% of Injuries** |
| --- | --- | --- |
| **FEMALE** | 23.8 | 30.9 |
| **MALE** | 23.3 | 69.1 |

After analyzing the data, it is clear that these cases are more accurately described as injuries due to people getting startled, or being around animals who did. This is supported by the randomness seen in the injury-rate by age beyond the teenage years.  
While horses are on top of the list of animals that caused injury by startling people or getting startled themselves, other notable ones include:

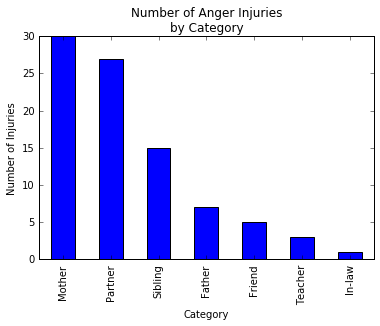
* Dogs: 5 incidents
* Cats: 2 incidents
* Friends: 1 incident
* Cousins: 1 incident
* Dragon flies: 1 incident (e.g. "6YF JUMPING ON TRAMPOLINE, GOT SCARED BY DRAGON FLY&FELL OFF LANDING ONTO SHOULDER>>HUMRUS FX"
* Fake Spider: 1 incident (e.g. "6YOF-PT FELL BACKWARDS INTO A GLASS DOOR AFTER BEING AFRAID OF A FAKESPIDER. GLASS BROKE BUT TO BUTTOCKS. DX- LACERATION TO LEFT BUTTOCKS”)

**Anger leads to…punching**

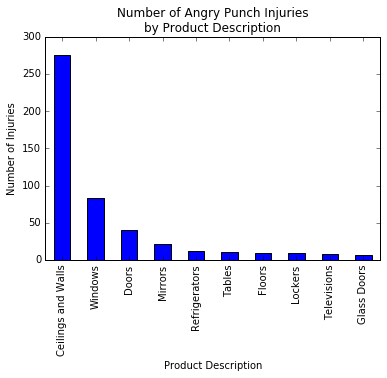
Anger returned 839 results, based on searching for "angry", "anger", "mad", "upset" or "frustrated". This is by far the emotion with the highest number of injuries related to it. Frustration was separated from anger as a separate emotion at the beginning of the analysis. However, after reading through the cases, the similarity between the two was high. Particularly in the presence and prevalence of wall-punching behavior in both categories.

The male of the species is more than twice as likely to injure himself due to anger than the female, as evidenced by the percentages of the injuries affecting each sex. The mean age for these injuries was similar for males and females, being in the mid twenties for both. When analyzed by age, a peak in injuries is seen in the teenage years for both sexes. Beyond that, a steady decline in injuries is seen as both males and females age.

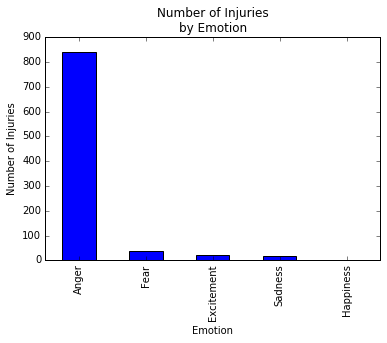
This could indicate that as people get older they get wiser and are less prone to anger, or that people find ways to channel their anger in a way that does not injure them.



The data was analyzed to find out whom the patient was angry with at the time of their injury, and sorted to find the most common people.  
Surprisingly, mothers top the list, with many people (average age of 14.3 years) hurting themselves due to being angry at their mothers. (e.g. 17 YOM HAD A FIGHT WITH BROTHER, MOM TOLD HIM TO LEAVE UNTIL HE "COOLEDOFF" - TRIED TO GET IN HOUSE, GOT MAD ,PUNCHED WINDOW. DX; HAND LAC).

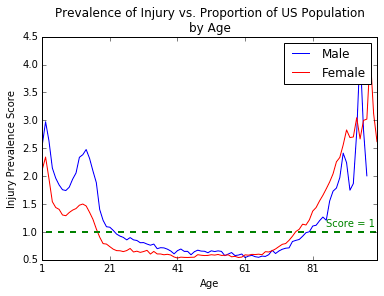
Coming in a close second is 'partner' which in this analysis is used to indicate a romantic partner. Siblings come in third, and fathers a distant fourth.  
It seems counterintuitive that the people thought to be closest to us are the ones that make us angry enough to hurt ourselves. It seems like the intensity of the emotional connection, not necessarily whether it is positive or negative, is what predicts the likelihood of injury.

One of the interesting observations on the the anger dataset, is the prevalence of people punching inanimate objects, mainly walls. This was exhibited in a remarkable 65% of the cases where anger was involved. As a point of comparison, kicking was exhibited in around 3% of the cases. The text search was carefully performed to exclude records of people getting punched or kicked, as opposed to doing the punching or kicking themselves. The demographics of the wall-punchers within the datasets closely follow the demographics of the Anger Injuries in general with the males being more than twice as likely to injure themselves as the females, and the mean age for both sexes is in the mid twenties.



**Injury Prevalence Score**

In an attempt to understand the risk of injury for various groups, the data was analyzed by age and sex, and compared to data from the 2015 US Census. One way to measure this risk is to compare the prevalence of injury in the data for each sex and age, and divide that by the proportion of the population that segment represented in the 2015 US Census data. The resulting score would be 1 if the prevalence of injury in the data was the same as the proportion of the population that segment represents. A score higher than 1 would indicate an over-representation in the data of the affected segment in relation to its proportion of the population (more injury-prone), and a score below one would represent the opposite (less injury-prone).



Based on the use of the Injury Prevalence Score (IPS), a few observations can be made. Firstly, for a long span of adulthood, the IPS is lower than 1. For males, this is between the ages of 23 and 79 and for females, it is between the ages of 18 and 75. This might point to a lower risk of injury at those ages.

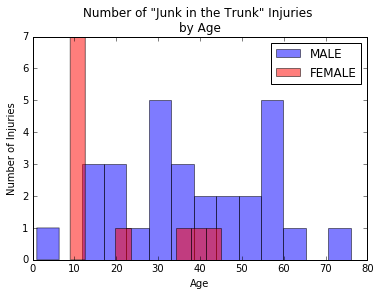
In the early years of life, the IPS is quite high, with a significant peak at age 2 for both sexes. Culturally, this might have been perceived hence the infamous "terrible twos". Another peak is seen at age 14 for both sexes as well. This might be due to the quick changes that occur around puberty, and the resulting clumsiness that some feel as they get accustomed to their growing bodies. Teens at this age might also be gaining more independence from their care givers, allowing for more opportunities for injury.

In the later years, starting at 75 for females and 78 for males, the IPS score is now greater than 1, and shows a steep increase by age for both ages.

Another clear trend is that females are less prone to injury than males, from birth through the age of 61, at which point women start to have a higher IPS score than men.

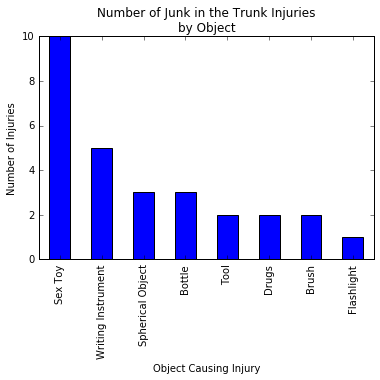
|  | **Mean age** | **% of Injuries** |
| --- | --- | --- |
| **FEMALE** | 19.7 | 27.5 |
| **MALE** | 37.9 | 72.5 |

## Junk in the Trunk

This search returned 40 results, using the search term “rectum".

Males are almost three times more likely to suffer this type of injury, as evidenced by the percentages of the injuries affecting each sex.  
The mean age for these injuries for males and females is quite different, with mean age for male patients being around double that of female patients. For females, 63% of the reported injuries occur in patients before their teens. The remaining 37% of cases are spread between the ages of 20 and 50. This implies that this behavior in females is largely restricted to children doing what children do.

In males, the cases reported show the behavior starting in the teen years. Unlike in the female population, the behavior continues throughout the whole lifespan. Other peaks are seen in the thirties and late fifties, with 79% of the cases reported in patients older than 20 years old.



The top two products used by the patients to cause this injury are vibrators/sex toys, used in 26% of the cases (e.g. 55YOM W/FOREIGN BODY IN RECTUM 2/2 DILDO IN RECTUM X 2 DAYS THAT HE WASUNABLE TO GET OUT), and pens/pecils used 13% of the cases (e.g. 22 YO MALE PLACED A BINGO DAUBER INTO HIS RECTUM. DX FOREIGN BODY RECTUM B) Other categories include balls/spherical objects used in 8% of the cases (e.g. 40 YR OLD MALE PUT BASEBALL UP RECTUM AND UNABLE TO REMOVE IT AND TAKENTO OR TO REMOVE), and bottles, also in 8% of the cases. Other notable objects are tools, brushes, and drugs (which were all related to smuggling, e.g. 20YOM HID PLASTIC BAGS OF MJ IN RECTUM WHEN ARRESTED & INGESTED THEM INJAIL TO AVOID MORE CHARGES. DX FB INGESTION, MJ INTOXICATION) used in 5% of the cases, respectively.

**Summary of recommendations**

From analysis of the NEISS data, we were able to form several conclusions as insight for our audience to lower their risk of injury. Our assessments were primarily related to analysis of sex and age, although consumer-products also played a large role.

1. Avoid fireworks on Fourth of July, footballs on Thanksgiving, knives in general.
2. Men should be making an effort to help women more in the kitchen on holidays so that they have their share of knife-related injuries.
3. When on a farm, avoid horses and fences.
4. Emotion, especially anger, plays a large role in injuries. People who are sad should stay away from floors, knives and walls. It is tough to be a mother, who are often injured by their angry teenage children.
5. To avoid injury, only orifice that should be receiving things is one’s mouth. For food.