

# Workplace Injuries in the United States 2015-2017

*Jordan Hunter*

*10/26/2017*

```
library(tidyverse)
library(tidytext)
library(readxl)
library(stringr)
library(sf)
library(ggribes)
library(RColorBrewer)
library(viridis)
library(scales)
library(forcats)
```

## Executive summary

My main questions are, “Where do workplace injuries occur, and what is the nature of these injuries?” I developed an interest in workplace safety when I studied work psychology in my undergrad. Workers can display interesting and surprising behavior while on the job, especially depending on their environment. The visuals I have prepared show a few main ideas: (1) Some regions of the United States, due to the prevalent industries in the area, have more injuries than others, (2) Certain body parts are more likely to be injured than others, and (3) the “Final Narrative” of an injury case reveals interesting information about what happens in certain industries.

## Data background

My data came from OSHA.gov, the website for the Occupational Safety and Health Administration. The data contains over 20,000 rows of data, each row a reported injury case from 2015 to 2017.

## Data cleaning

I had to manipulate data on industry names and on the body parts that were injured. For certain information to be mapped, I had to tidy-up the data into summary tables, often showing data by groups.

## Individual figures

Describe and show how you created the first figure. Why did you choose this figure type? What did you do to ensure it follows Cairo’s principles from *A Truthful Art* and Robin Williams’s CRAP?

I’ve never seen a map that shows where workplace injuries occur in the United States. I thought it would be interesting to visualize this. I was worried it might just be a population map - in some ways it is, but really it’s not. Otherwise, California and Arizona should have way more going on. I used the yellow fill color and black points - which contrast nicely - to resemble caution tape (these colors are also on OSHA’s website). I used the Albers projection of the United States.

Where Do Workplace Injuries Get Reported to OSHA?

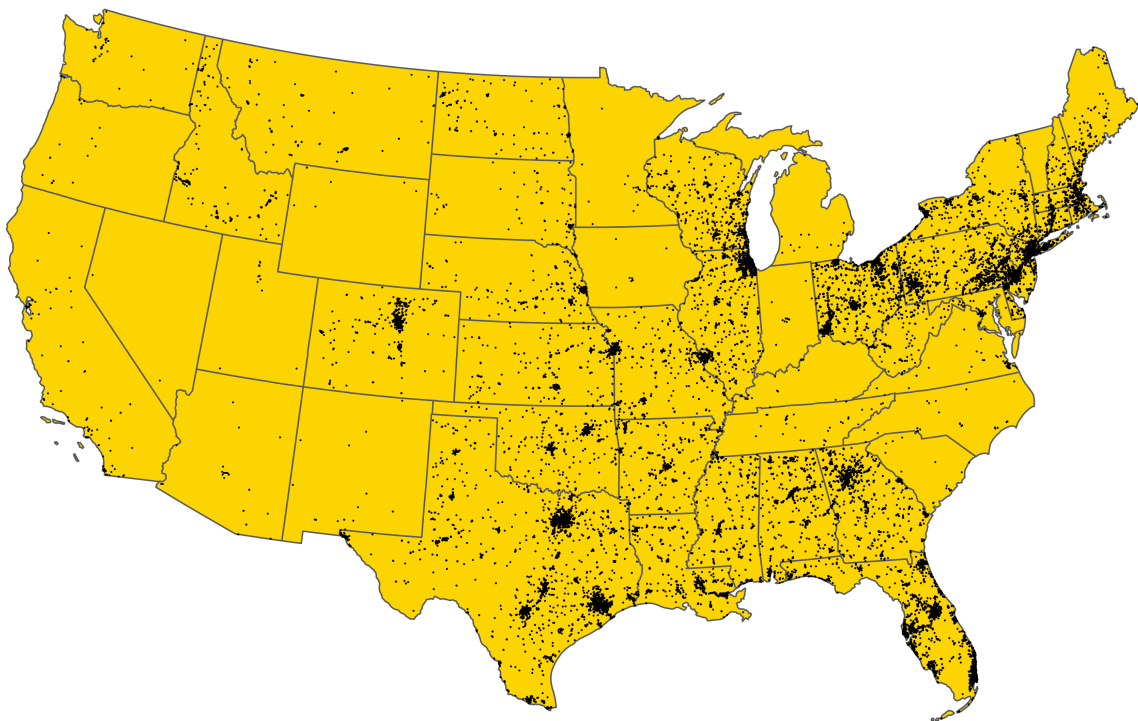


Figure 1: Workplace Injuries in The United States

```

injuries_raw <- read_csv("Data/severeinjury.csv")
naics <- read_xls("data/2-digit-naics-sector-codes.xls") %>%
  rename("NAICS.Sector" = `NAICS Sector`, "NAICS.US.Title" = `NAICS US Title`)

naics <- read_xls("Data/2-digit-naics-sector-codes.xls") %>%
  rename("NAICS.Sector" = `NAICS Sector`, "NAICS.US.Title" = `NAICS US Title`)

leftout <- data.frame("NAICS Sector" = c(31, 32, 33, 44, 45, 48, 49),
  "NAICS US Title" = c(rep("Manufacturing", 3),
    rep("Retail Trade", 2),
    rep("Transportation and Warehousing", 2)))

naics <- rbind(naics, leftout)

injuries_clean <- injuries_raw %>%
  mutate(industryid = as.numeric(substr(`Primary NAICS`, 1, 2))) %>%
  left_join(naics, by = c("industryid" = "NAICS.Sector")) %>%
  rename("Industry.Name" = NAICS.US.Title) %>%
  filter(!is.na(Longitude)) %>%
  st_as_sf(coords = c("Longitude", "Latitude"), crs = 4326)

state <- st_read("Data/cb_2016_us_state_20m/cb_2016_us_state_20m.shp",
  stringsAsFactors = FALSE) %>%
  mutate(State = str_to_upper(NAME))

## Reading layer `cb_2016_us_state_20m' from data source `~/Users/Master/Documents/MPA/MPA Fall 2017/Data'
## Simple feature collection with 52 features and 9 fields
## geometry type: MULTIPOLYGON
## dimension: XY
## bbox: xmin: -179.1743 ymin: 17.91377 xmax: 179.7739 ymax: 71.35256
## epsg (SRID): 4269
## proj4string: +proj=longlat +datum=NAD83 +no_defs

stateclean <- state %>%
  filter(!(STUSPS %in% c("AS", "PR", "GU", "MP", "VI")))

stateclean_48 <- state %>%
  filter(!(STUSPS %in% c("AS", "PR", "GU", "MP", "VI", "AK", "HI")))

Lat1 <- 50.816348
Lat2 <- 22.257326
Lon1 <- -63.193359
Lon2 <- -128.935547

injuries_transformed_48 <- cbind(injuries_clean, st_coordinates(injuries_clean)) %>%
  st_transform(crs = 102003) %>%
  filter(Y < Lat1, Y > Lat2, X < Lon1, X > Lon2) %>%
  select(-X, -Y)

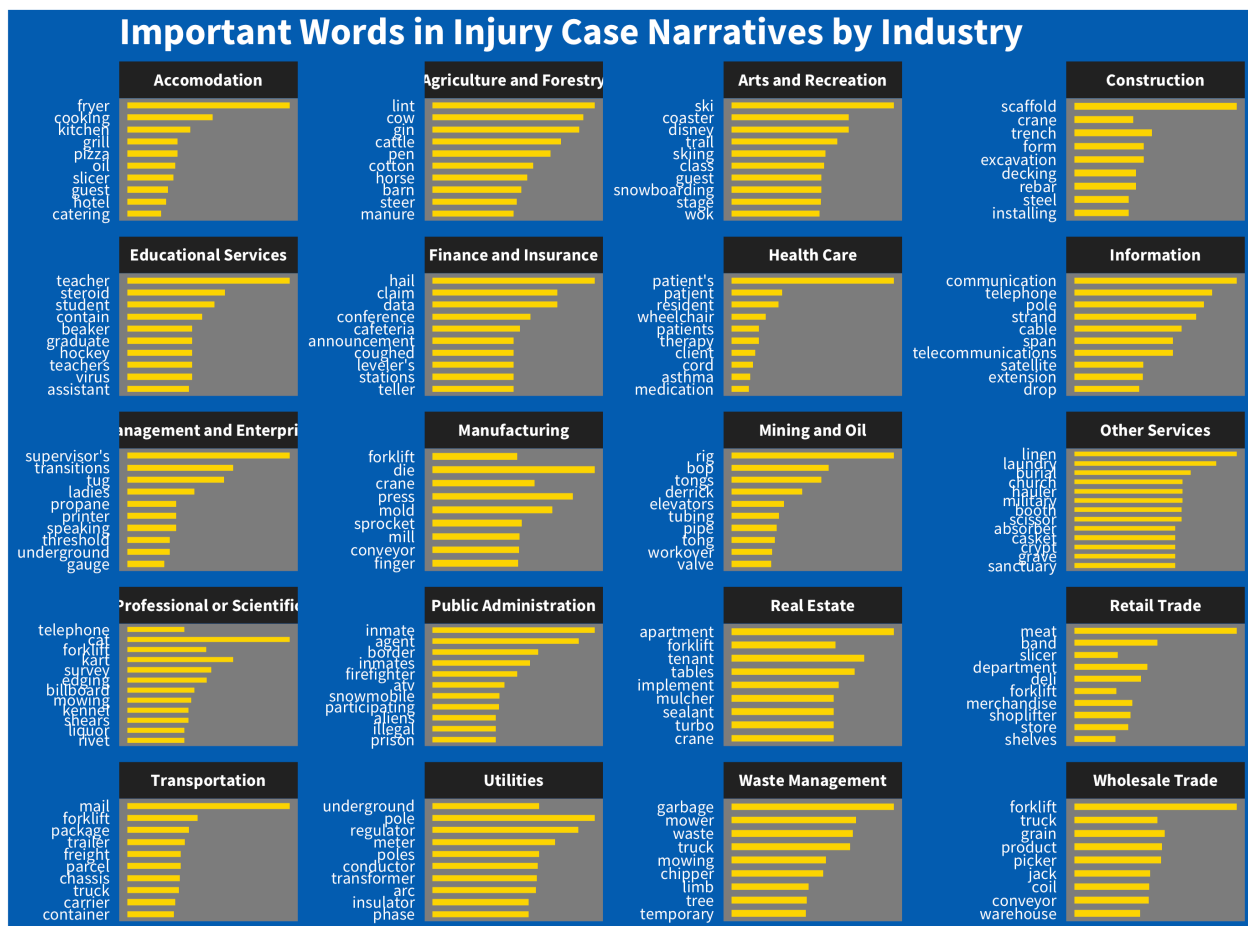
injuries_transformed_48_forreal <- cbind(injuries_transformed_48,
  st_coordinates(injuries_transformed_48))

US_injuries <- ggplot() +
  geom_sf(data = stateclean_48, fill = "gold") +

```

```
geom_point(data = injuries_transformed_48_forreal,
           aes(x = X, y = Y), size = .01) +
coord_sf(crs = 102003, datum = NA) +
theme_void() +
theme(panel.background = element_rect(fill = "transparent", colour = NA),
      plot.background = element_rect(fill = "transparent", colour = NA),
      panel.grid.major.x = element_blank(),
      panel.grid.major.y = element_blank(),
      panel.grid.minor = element_blank(),
      axis.ticks = element_blank())
```

Figure 2



These colors, again from OSHA's website, have great contrast. I also keep everything left-aligne (except for the panel titles). This visual is my favorite, because the information is really fascinating. Some insights are also hilarious. Apparently, there is some guy named "derrick" that is strongly related to injuries that occur in the Mining and Oil industries. *Note: My code is for a condensed version, which I used in my final product. I included the more granular visual here because it is more fun to read*

```
injured_part <- data_frame(body_part = unique(injuries_clean$`Part of Body Title`))

#write_csv(injured_part, "Output/injured_part.csv")
```

```

injured_part_agg <- read_csv("Output/injured_part.csv")

injuries_clean_agg <- injured_part_agg %>%
  left_join(injuries_clean) %>%
  separate_rows(body_part, sep = ",") %>%
  mutate(body_part = str_trim(body_part))

body_part <- injuries_clean_agg %>%
  mutate(Industry.Name = recode(Industry.Name,
                                `Wholesale Trade` = "Merchandise",
                                `Retail Trade` = "Merchandise",
                                `Transportation and Warehousing` = "Transportation",
                                `Administrative and Support and Waste Management and Remediation Services` = "Waste Management")) %>%

  group_by(Industry.Name, body_part) %>%
  summarize(total = n()) %>%
  ungroup() %>%
  filter(!is.na(Industry.Name)) %>%
  arrange(desc(total)) %>%
  filter(Industry.Name %in% c("Manufacturing",
                             "Construction",
                             "Merchandise",
                             "Transportation",
                             "Waste Management")) %>%
  mutate(body_part = fct_rev(factor(body_part, ordered = TRUE, levels = c("head",
                                                                           "arms",
                                                                           "trunk",
                                                                           "hands",
                                                                           "groin/hips",
                                                                           "legs",
                                                                           "feet",
                                                                           "whole body"))),
         Industry.Name = fct_inorder(Industry.Name, ordered = TRUE))

bodyparts_injured_plot <- ggplot(body_part, aes(x = body_part, y = total)) +
  geom_pointrange(aes(ymin = 0, ymax = total, color = "gold")) +
  scale_color_manual(values = "gold") +
  coord_flip() +
  labs(y = "Injuries", x = NULL,
       title = "Hands Are More At Risk",
       subtitle = "Injured Body Parts In The Industries Most Likely to Report Severe Workplace Injuries") +
  scale_y_continuous(label = comma) +
  theme_dark(base_family = "Source Sans Pro") +
  theme(legend.position = "none",
        panel.grid.minor.x = element_blank(),
        panel.grid.major.y = element_blank(),
        strip.text = element_text(color = "white", face = "bold", size = 12),
        plot.title = element_text(size = 20, color = "white", face = "bold"),
        plot.subtitle = element_text(color = "white", size = 14),
        axis.text.x = element_text(size = 10, color = "white"),
        axis.text.y = element_text(size = 12, color = "white"),
        axis.title.x = element_text(size = 13, color = "white"),

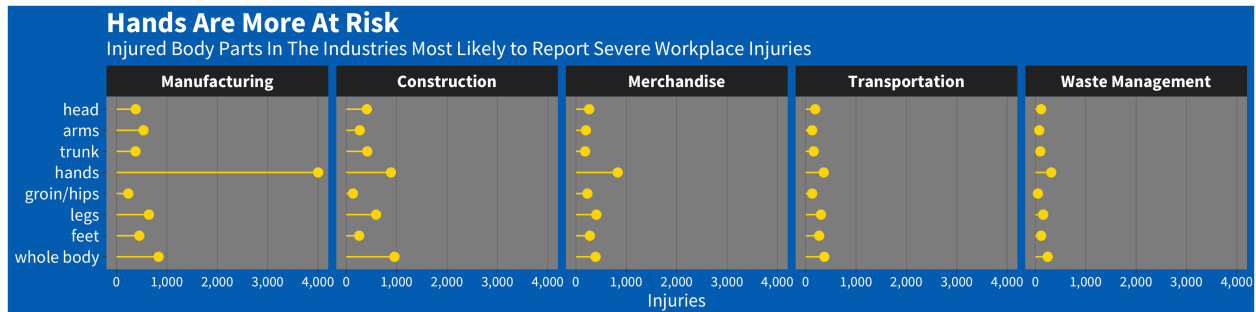
```

```

plot.background = element_rect(fill = "#025FB3"),
axis.text = element_text(color = "black")) +
facet_wrap(~ Industry.Name, nrow = 1)

```

Figure 3



This visual is fun. It shows which body parts are injured the most in the top 5 most injury-prone industries.

```

narrative <- data_frame(narrative = injuries_clean$`Final Narrative`, industry = injuries_clean$Industry)
mutate(industry.name = recode(industry,
  `Transportation and Warehousing` = "Transportation",
  `Mining, Quarrying, and Oil and Gas Extraction` = "Mining and Oil",
  `Accommodation and Food Services` = "Accommodation",
  `Administrative and Support and Waste Management and Remediation Services` = "Waste Management",
  `Health Care and Social Assistance` = "Health Care",
  `Agriculture, Forestry, Fishing and Hunting` = "Agriculture and Forestry",
  `Arts, Entertainment, and Recreation` = "Arts and Recreation",
  `Other Services (except Public Administration)` = "Other Services",
  `Professional, Scientific, and Technical Services` = "Professional or Scientific",
  `Management of Companies and Enterprises` = "Management and Enterprise",
  `Real Estate and Rental and Leasing` = "Real Estate"))

tidy_narrative <- narrative %>%
  unnest_tokens(word, narrative) %>%
  count(industry.name, word, sort = TRUE)

narrative_unique <- tidy_narrative %>%
  bind_tf_idf(word, industry.name, n)

narrative_top_10 <- narrative_unique %>%
  group_by(industry.name) %>%
  top_n(4, tf_idf) %>%
  ungroup() %>%
  filter(!(word %in% "he")) %>%
  arrange(desc(tf_idf)) %>%
  mutate(word = fct_rev(fct_inorder(word, ordered = TRUE)))

injury_narrative_plot <- ggplot(narrative_top_10, aes(x = word, y = tf_idf)) +
  geom_col(width = .4, fill = "gold") +
  coord_flip() +
  labs(y = NULL, x = NULL, title = "Important Words in Injury Case Narratives by Industry") +
  theme_dark(base_family = "Source Sans Pro") +

```

```

theme(axis.text.x = element_blank(),
      panel.grid = element_blank(),
      axis.ticks = element_blank(),
      plot.title = element_text(size = 20, face = "bold", color = "white"),
      plot.background = element_rect(fill = "#025FB3"),
      axis.text = element_text(color = "white"),
      strip.text = element_text(face = "bold", color = "white", size = 8)) +
guides(fill = FALSE) +
facet_wrap(~ industry.name, scales = "free", nrow = 5)

```

## Final figure

I figured some stuff out and I feel really good about my final product. It may not be the most visually stunning graphic, but it is simple and can share a huge amount of information about the nature of workplace injuries in the United States. I used the same color scheme used by OSHA.

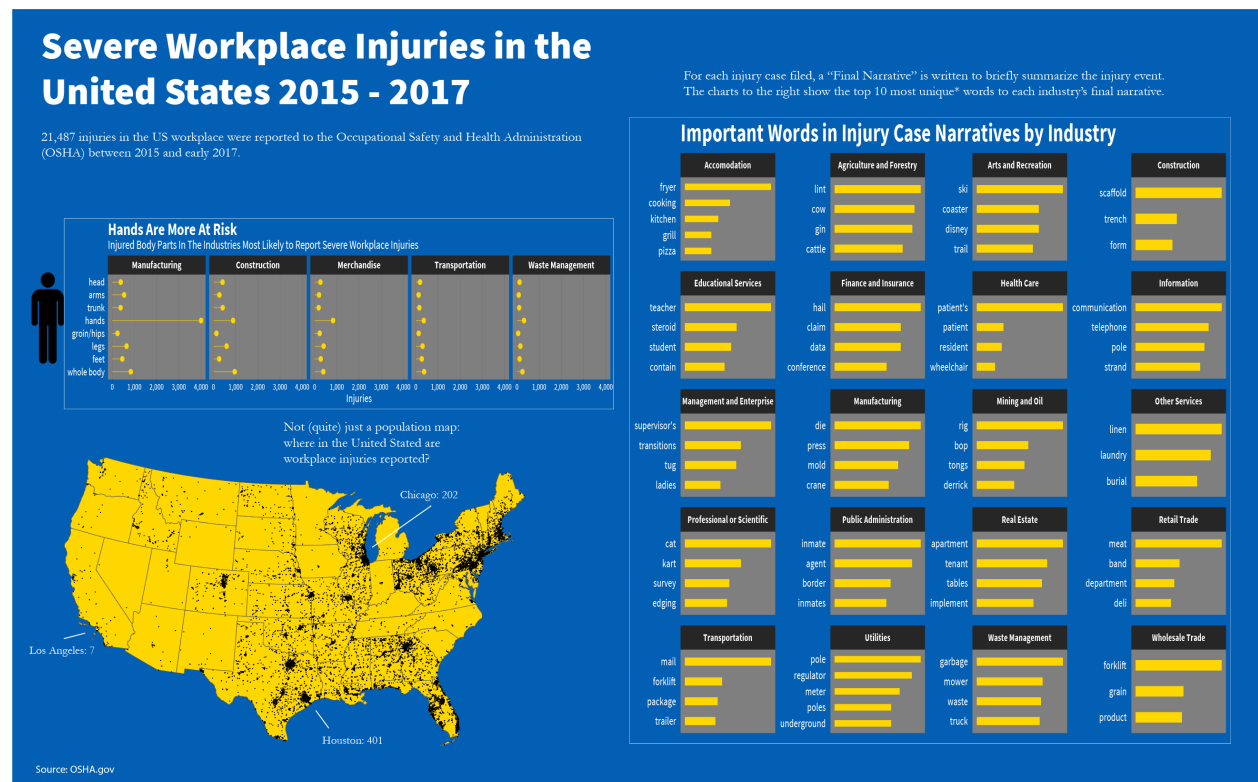


Figure 2: Workplace Injuries