Work-Related Injuries & Deaths 2018-2020

The main goal of this project was to find out which states were the least and most dangerous to work in and to find out which industries were the least and most dangerous to work in.

For this project I downloaded the workplace injury data in .csv format (ITA Data CY 2018.csv, ITA Data CY 2019.csv, ITA Data CY 2020.csv) from
https://www.osha.gov/Establishment-Specific-Injury-and-Illness-Data

I loaded those three into Power BI and removed some of the unnecessary columns so I would have less cluttered data to work with. I then appended these three tables into a new, single table for use in my reports.

At this point I went through my combined table and removed more unneeded columns. The required columns to create my report were more obvious at this point, making this a good time to do this.

I removed multiple rows from the table with no data in them.

Many of the text columns (city, state, company name, establishment name, and industry description) were inconsistently formatted, so I then took care of as much of that as I reasonably could. A lot of this was capitalizing each word and trimming spaces from the ends.

For entries including the words “Inc”, “Co”, and “Llc” the formatting was especially troublesome, because they would sometimes come after a comma and sometimes not. They would sometimes have a space between them and the comma, sometimes not. I replaced the versions of those including the comma with the non-comma version, then replaced the entire thing with a “, Inc”, etc.

There were multiple rows with no company name. For these, I copied the establishment name into that field so we would have a company name to work with.

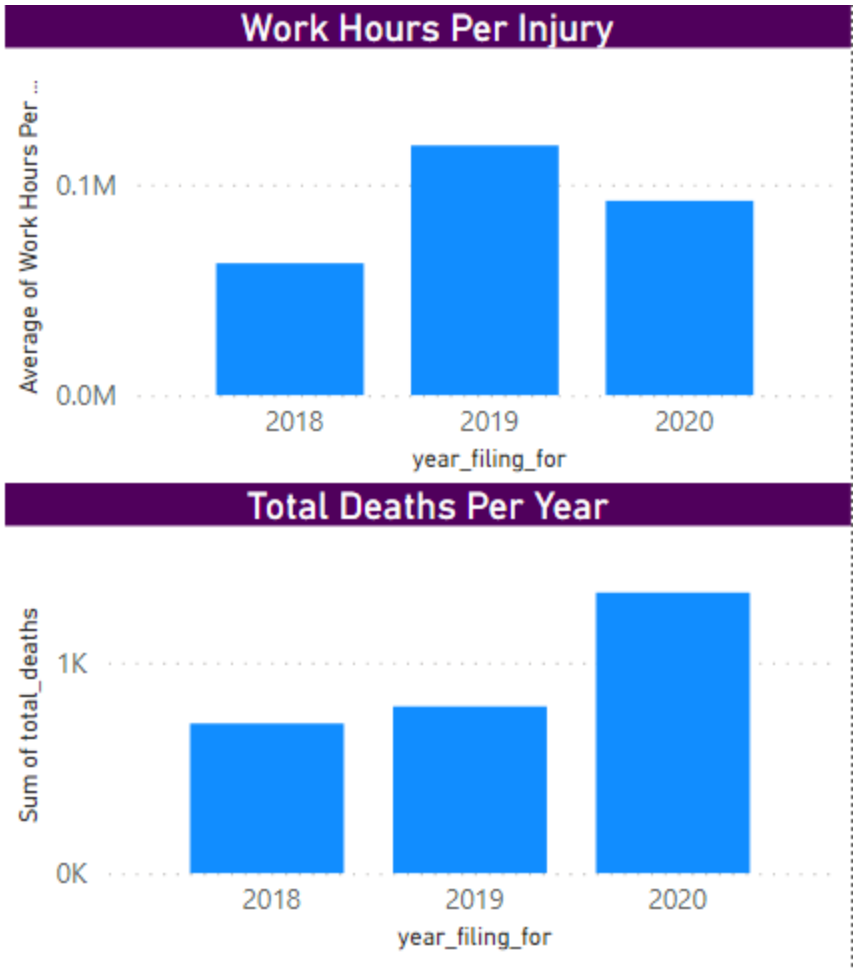
When glancing through the data in a matrix view, I noticed a few instances of states that didn’t make sense to me. For example, there was an entry with a state “FM”. I researched to make sure this wasn’t an established and known postal abbreviation (such as AP being a code for “Armed Forced Pacific”), then looked up the rest of the address and establishment at the address without a state listed. In those cases, I was able to identify the actual state (in that case it was Florida) and correct the data in the table manually. This was a small number, so manual adjustment seemed to be the proper way to correct the issue.

At this point I turned to the reports.

My first thought was to show the the total injuries, work hours per injury, and deaths per state during the 2018-2020 timeframe. I added a matrix to show this data. In this matrix you can drill down to city as well, if you want to find out which part of the state had the highest number of injury and death reports.

state	Injuries	Work Hrs/ Injury	Deaths
WY	4,120	217,639.52	4
DC	5,283	44,174.46	7
AK	7,645	25,499.08	4
VT	7,862	30,960.79	2
MT	8,469	31,588.85	3
DE	8,845	49,120.71	8
HI	9,885	35,598.58	3
RI	11,564	46,937.62	10
ND	11,793	45,283.01	6
NM	12,546	47,260.23	16
WV	13,116	42,095.99	14
SD	13,491	294,315.62	13
NH	17,313	47,811.28	7
ME	19,879	29,995.64	9
ID	21,262	42,500.90	14
NE	23,204	44,540.17	26
AR	24,071	82,366.11	63
LA	24,927	75,964.33	43
MS	26,222	96,730.62	30
UT	28,903	180,332.19	25
KS	31,307	41,061.01	30
OK	32,295	49,988.23	48
NV	42,659	88,498.87	24
AL	42,752	85,204.55	58
MD	45,039	58,311.34	33
SC	47,203	54,879.27	54
IA	52,730	53,560.30	47
CT	56,293	52,907.54	25
CO	57,884	59,625.37	41
OR	58,046	101,844.46	34
KY	59,437	48,145.15	62
MO	63,337	54,599.34	57
Total	**3,372,966**	**92,125.32**	**2,836**

It also made sense to show the average work hours per injury for each year and deaths for each year, so I added a stacked column chart for each of these.



I then added cards with the overall total injuries and the total deaths to make these number apparent.

—

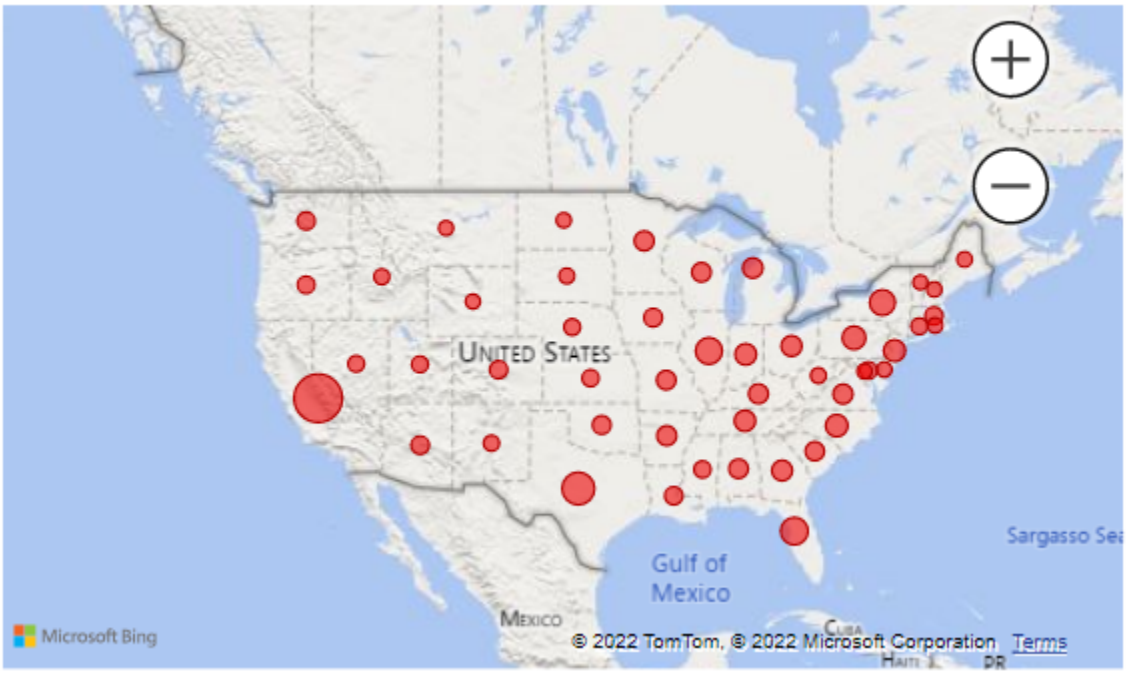
It made sense to get state by state breakdown, so I added a tab to show that. On that tab there is a similar matrix as on the first page, but you can't drill down to city because this tab specifically looks on a state by state basis.

State	Total Injuries	Work Hrs/Injury	Total Deaths
CA	440,612	208,045.01	400
TX	211,216	123,238.09	215
FL	182,449	56,476.35	154
IL	155,951	58,262.20	145
NY	139,826	54,985.57	132
PA	171,102	187,976.28	106
NC	102,159	52,413.22	92
NJ	74,287	56,253.48	86
IN	94,357	58,869.88	82
TN	79,235	80,144.44	79
OH	133,258	61,095.52	74
GA	95,504	78,330.74	73
MI	113,988	73,646.13	68
AR	24,071	82,366.11	63
MN	98,178	47,428.12	63
VA	82,587	57,365.95	63
KY	59,437	48,145.15	62
AL	42,752	85,204.55	58
WI	90,956	52,356.32	58
MO	63,337	54,599.34	57
SC	47,203	54,879.27	54
OK	32,295	49,988.23	48
IA	52,730	53,560.30	47
AZ	69,163	192,422.39	44
LA	24,927	75,964.33	43
CO	57,884	59,625.37	41
MA	66,097	56,902.06	41
WA	82,659	105,269.04	41
OR	58,046	101,844.46	34
MD	45,039	58,311.34	33
KS	31,307	41,061.01	30
MS	26,222	96,730.62	30
Total	**3,372,966**	**92,125.32**	**2,836**

This page also contains a bubble map that shows a larger bubble for states with more injuries and one with a larger bubble for states with more deaths.

Avg. Hours Per Reported Work-Related Injuries

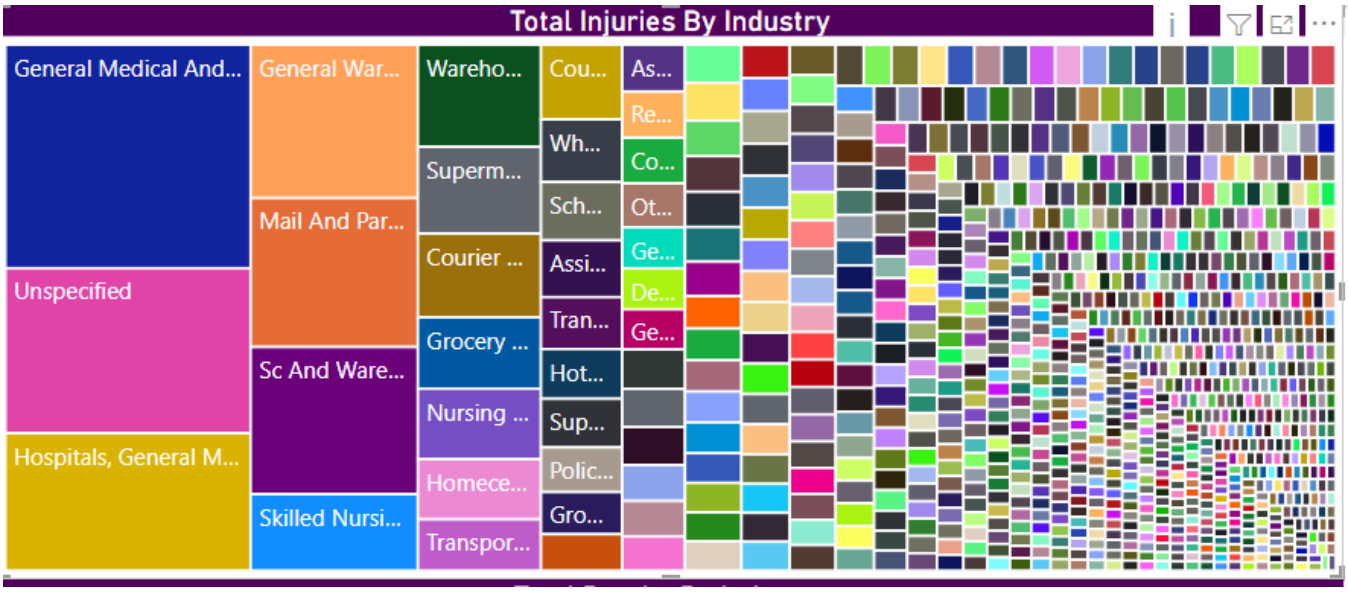
Reported Work-Related Deaths



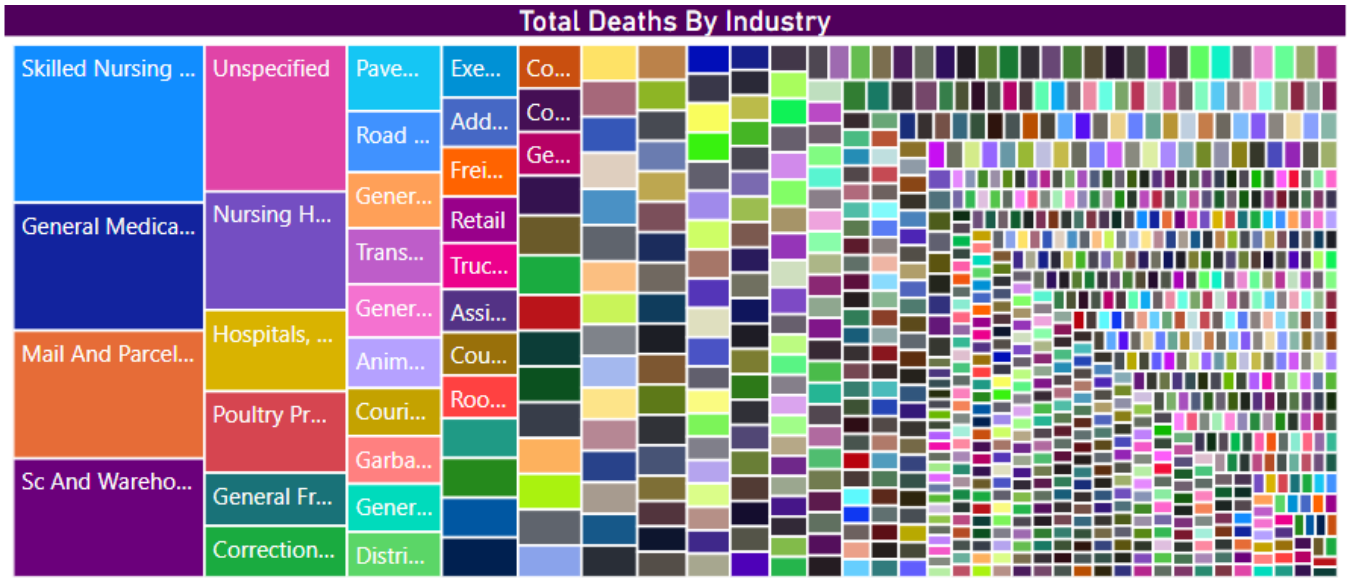
I added a tab to show the injury and death data per industry. It made sense to put that into a matrix, so I did that. To add more impact, I put color bars on the injuries to highlight the industries with the highest number of injuries. I also set it so that the background on the total deaths column are darker red for industries with more deaths to make that stand out.

Industry	Total Injuries	Total Deaths
Skilled Nursing Facilities	49,933	120
General Medical And Surgical Hospitals	214,442	97
Mail And Parcel Delivery	97,567	97
Sc And Warehouse	96,332	90
Unspecified	158,414	83
Nursing Homes	33,984	67
Hospitals, General Medical And Surgical	131,605	46
Poultry Processing	4,003	46
General Freight Trucking, Long-Distance, Truckload (Tl)	7,669	30
Correctional Institutions	11,235	29
Pavement, Highway, Road, Street, Bridge Or Airport Runway, Construction	5,483	25
Road Construction	3,975	23
General Warehousing And Storage	100,352	21
Transportation Air Cargo	24,651	21
General Services Departments, Government	8,396	20
Animal (Except Poultry) Slaughtering	2,743	19
Couriers	23,785	18
Garbage Collection Services	5,057	18
General Freight Trucking, Long-Distance, Less-Than-Truckload (Ltl)	10,090	18
Distribution Of Electric Power	7,881	17
Executive And Legislative Office Combinations	6,804	16
Addition, Alteration And Renovation General Contractors, Commercial And Institutional Building	3,266	15
Freight Trucking Ltl	7,174	15
Retail	7,609	14
Trucking, General Freight, Long-Distance, Truckload (Tl)	4,690	14
Assisted-Living Facilities With On-Site	11,493	13
Total	**3,372,981**	**2,836**

I then added a treemap chart for total injuries by industry so we could see which industry had the most injuries.



I also added a treemap for total deaths for each industry to show which industry leads to the most deaths.



It was at this point that I realized I had a lot of blank entries for the industry. I went back into the Power BI Query Editor and changed those to “Unspecified”. It seemed reasonable, given the fact that there was no industry specified at all.

I then created a tab with questions I think could be asked of this data, with a bookmark link back to the proper page and filter settings to answer that question.

This leads us to these questions and answers, which wouldn't necessarily have been obvious in the raw data:

- Which state has the fewest total work-related injuries?
  Wyoming, with 4,120 from 2018-2020
- Which state has the highest number of work hour per injury, making it least dangerous from this point of view?
  California, with an average of over 208,000 hours per injury
- Which state has the least work-related deaths?
  Vermont, with only 2 during this timeframe
- Which state has the most work-related injuries?
  California, with over 440,000
- Which state has the lowest number of work hours per injury, making it most dangerous from this point of view?
  Vermont, with 7862
- Which state has the most work-related deaths?
  California, with 400 during this timeframe
- Which industry has the most work-related injuries?
  General Medical and Surgical Hospitals, with over 214,000 injuries during the 2018-2020 timespan
- Which industry has the most work-related deaths?
  Skilled Nursing Facilities, with 120 in the 2018-2020 years

Something that becomes apparent from these questions and answers is that state size makes little difference to how dangerous a job is, it mostly comes down to industry. While I'm not sure most people would guess it, medical and nursing-related industries are some of the most dangerous jobs for both injuries and deaths during that timespan, having three of the top ten industries with most injuries and four of the top ten for deaths during the 2018-2020 timeframe.