All Deaths:

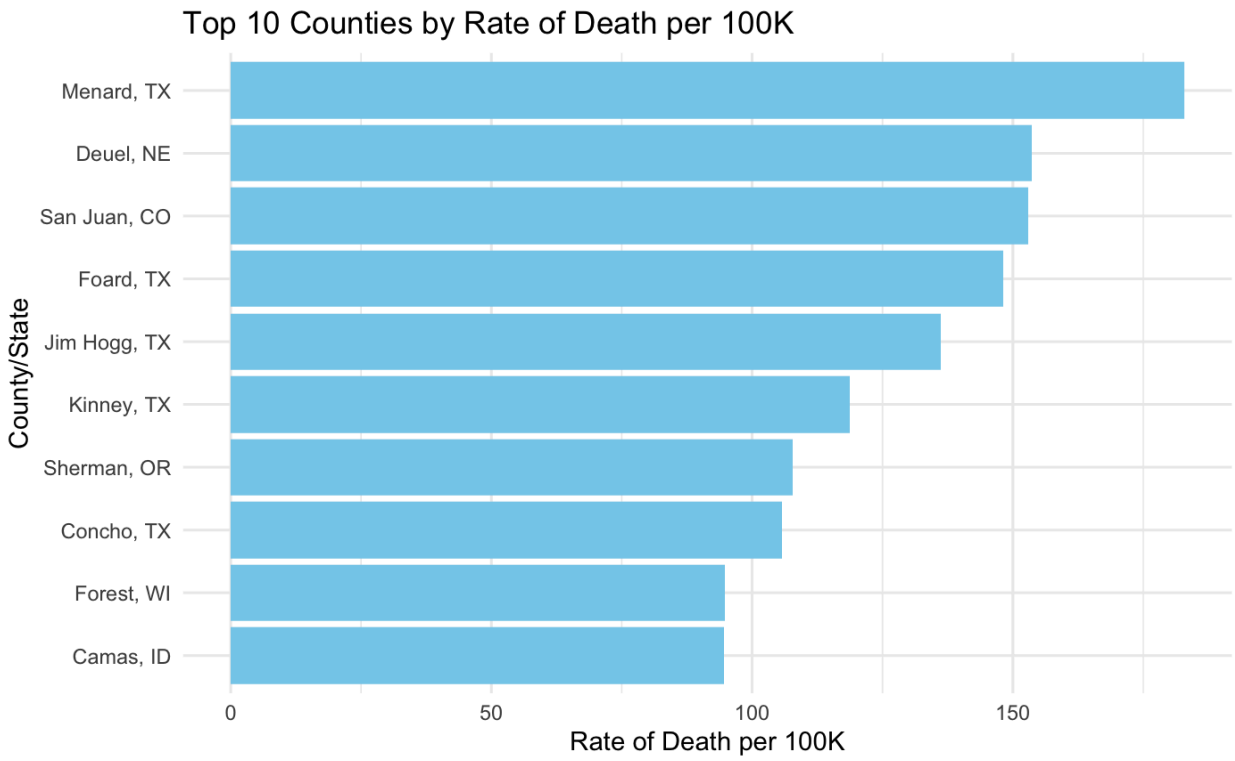
Numerical (does not take into account county populations):

countystate <chr>	n <int>	percentage <dbl>
Los Angeles, CA	1545	4.81248443
Cook, IL	640	1.99352106
Harris, TX	615	1.91564914
Maricopa, AZ	518	1.61350611
San Bernardino, CA	379	1.18053825
San Diego, CA	361	1.12447047
Riverside, CA	315	0.98118615
St. Louis, MO	311	0.96872664
Clark, NV	306	0.95315226
Orange, CA	298	0.92823324

Table (takes county populations into account):

County	Deaths/100k Residents
Menard, TX	182.87108
Deuel, NE	153.63605
San Juan, CO	152.98317
Foard, TX	148.07502
Jim Hogg, TX	136.20444
Kinney, TX	118.75309
Sherman, OR	107.83609
Concho, TX	105.69893
Forest, WI	94.70361
Camas, ID	94.57755

Bar chart (takes county populations into account):



GSW Deaths:

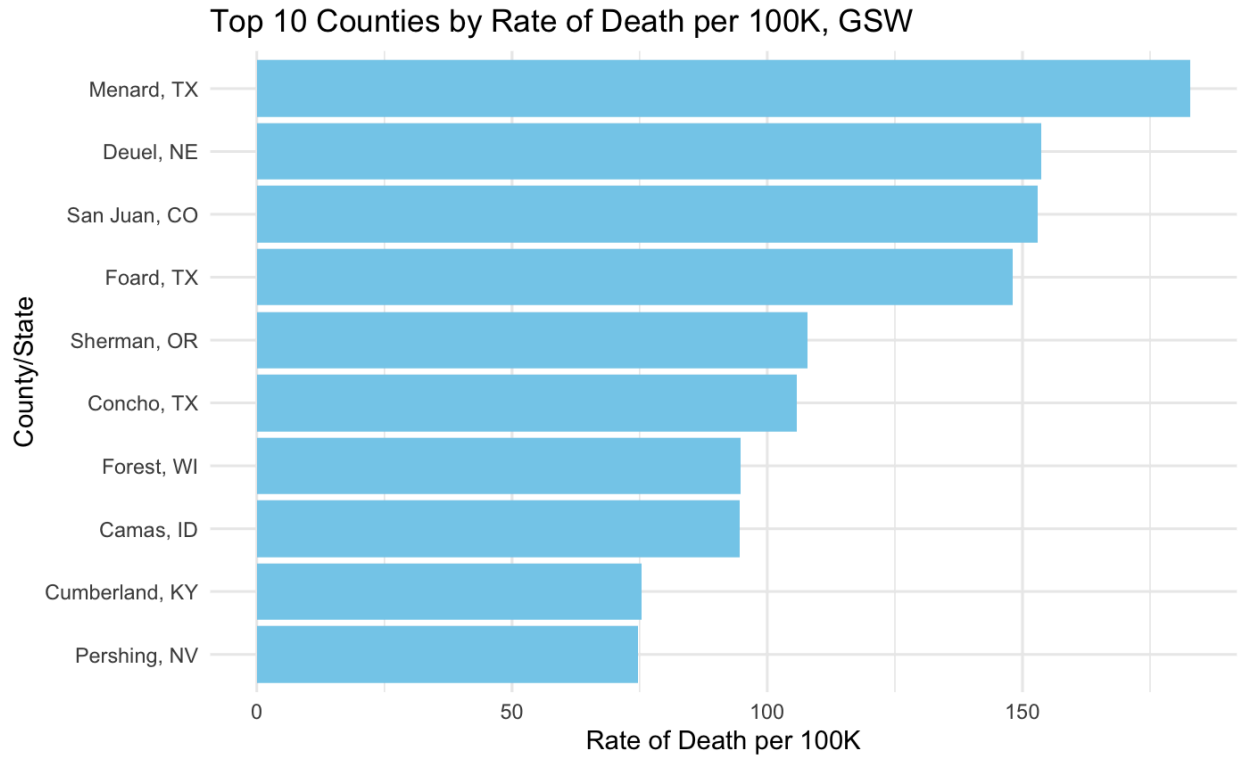
Numerical (does not take county population into account):

countystate <chr>	n <int>	percentage <dbl>
Los Angeles, CA	1265	5.43268198
Maricopa, AZ	465	1.99699377
Harris, TX	455	1.95404767
Cook, IL	407	1.74790638
San Bernardino, CA	297	1.27549925
Clark, NV	273	1.17242860
Riverside, CA	254	1.09083101
San Diego, CA	236	1.01352802
Orange, CA	223	0.95769809
Philadelphia, PA	209	0.89757355

Table (takes county populations into account):

County	GSW Deaths/100k Residents
Menard, TX	182.87108
Deuel, NE	153.63605
San Juan, CO	152.98317
Foard, TX	148.07502
Sherman, OR	107.83609
Concho, TX	105.69893
Forest, WI	94.70361
Camas, ID	94.57755
Cumberland, KY	75.45651
Pershing, NV	74.60831

Bar chart (takes county populations into account):



Non-GSW Deaths:

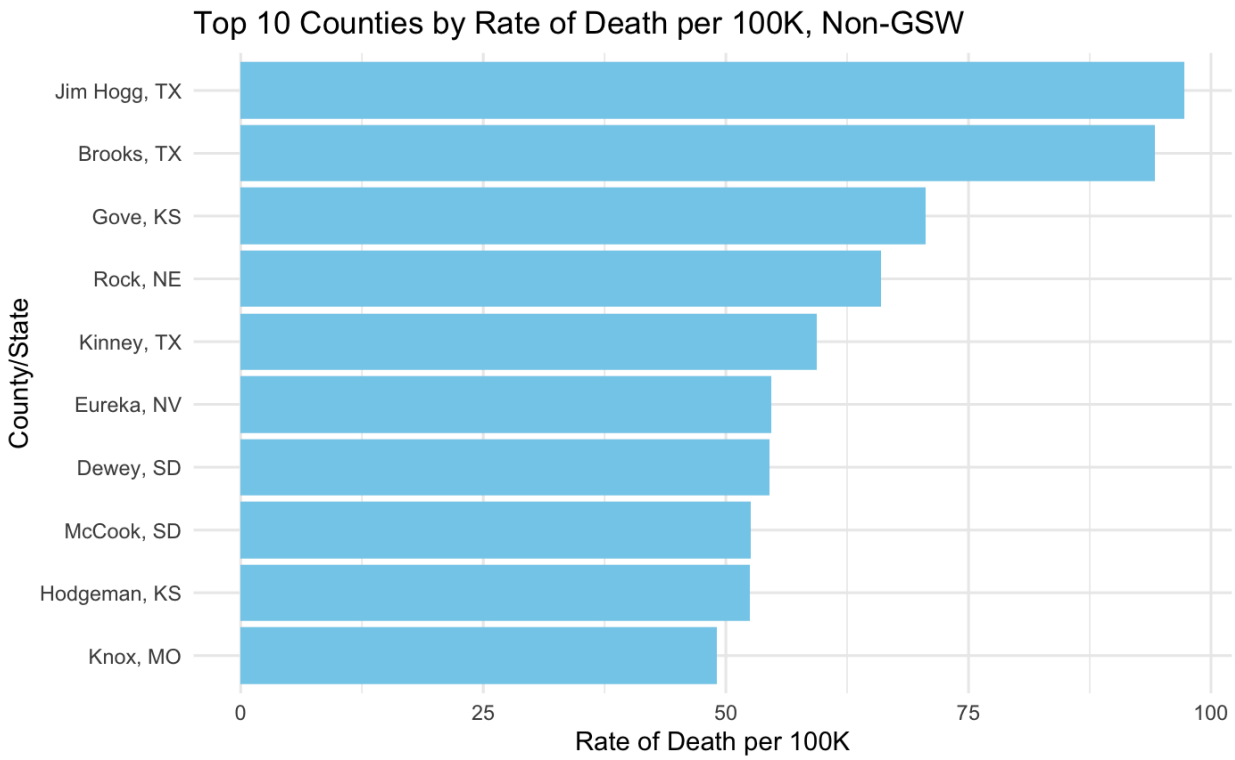
Numerical (does not take county populations into account):

countystate <chr>	n <int>	percentage <dbl>
Los Angeles, CA	280	3.17496315
Cook, IL	233	2.64202291
Harris, TX	160	1.81426466
Wayne, MI	126	1.42873342
San Diego, CA	125	1.41739426
St. Louis, MO	119	1.34935934
Bexar, TX	92	1.04320218
Alameda, CA	91	1.03186302
Dallas, TX	88	0.99784556
San Bernardino, CA	82	0.92981064

Table (takes county populations into account):

County	Non-GSW Deaths/100k Residents
Jim Hogg, TX	97.28888
Brooks, TX	94.22956
Gove, KS	70.61316
Rock, NE	66.00660
Kinney, TX	59.37655
Eureka, NV	54.64481
Dewey, SD	54.49261
McCook, SD	52.56395
Hodgeman, KS	52.44755
Knox, MO	49.04365

Bar chart (takes county populations into account):



Methodology:

I had the death data with county and state included, and I created a new column that combines county name and state into one value. The values in this column ("countystate") appeared like this: "Lake, IL", "Kern, CA", etc.
I removed the Year column, so my dataset only had the columns State, County, cause (cause of death), and countystate.

Numerical Data Methods:

(How I found number of deaths in each county)
I ran code that summed up the number of times the same value in the column countystate showed up in the dataset. In other words, since each row of the dataset represents an individual death, you add up the number of rows where "Lake, IL" was the county the victim passed. This then equals the number of deaths that occurred in Lake, IL, and those values were assigned a new column.

(How I filtered them out by cause of death)
To filter this out by whether the victims passed from gunshot wounds or other causes, I created two separate datasets. I created code that searched the dataset to find rows where the cause of death included the word "gunshot", and then created a dataset that is of victims who passed from gunshot wounds. I then did the opposite to create a dataset where the victims passed from any other cause of death other than gunshot wounds. For both datasets, I then ran the same code from

the paragraph before that adds up the total number of deaths in each county, and assigned those values to the column “n”.

Proportional Data Methods:

Datasets used to calculate county population are from [Index of /programs-surveys/popest/datasets]

(How I found the population for each county)
To find county population data, I took datasets from 2000-2010, 2010-2020, and 2020-2023.I standardized state names to state abbreviations, and created the column “countystate” once more. The census datasets were quite large, so I extracted the columns “ESTIMATESBASE20***XX***”, which is the census’ base estimate for the counties’ population. I then joined together all three datasets of census data by the countystate row. This then created a dataset where there was the countystate, population estimates from 2000, population estimates from 2010, and population estimates from 2020. I took the average of the three population estimates columns to create a new column called population.

(How I found the death rate per 100k residents)
I joined the death data (the dataset that was used for numerical calculations) and the population data by the countystate column. I then created a new column called “count_per_population”. The formula for that column was:

(Number of deaths occurred/average population)*100,000

This calculates the rate of deaths occurred per 100,000 residents of each county. I extracted the top 10 counties, and turned them into tables and bar charts. For the GSW and non-GSW deaths, I joined together their respective datasets (deaths only by GSW and deaths only by non-GSW) with the county population data.

Error I made originally (all images are updated with correct data):
Some of the sums for deaths in GSW/non-GSW were under/overcounted due to how I dealt with duplicate values. Originally, I recalculated GSW and non-GSW deaths for population comparisons and kept only distinct values for countystate (i.e. only kept one value for Lake, IL) that had incorrect counts for deaths. The new values should have the correct counts of GSW and non-GSW deaths, and then had them divided by population (then multiplied by 100k) to get deaths per 100k residents.