1. **Intro**
   1. About the data, why did we choose it, what are the questions we hope to answer, what did we find > business questions
   2. *Business questions*: Gun VS Vehicle deaths, How school affects lifespan, Cremation vs Burial, Deaths by Health/lifestyle conditions
   3. **From the Guidelines PPT**: [Provide answers to the following questions:

We are interested in better understanding gun-related deaths – where do they occur? When do they occur? What are correlating factors when guns or shootings are implicated?

Are there systematic patterns across sociodemographic variables in terms of the reasons for death?

We want to understand choices that people make between cremation and burial – can you develop a prediction model for whether a person is cremated or buried?]

1. **Descriptive analytics:**
   1. Put the dataset into MySQL
2. **Predictive analytics:**
   1. Get Spark started

Business Questions (links)

1. Guns vs Vehicle Deaths

<https://www.kaggle.com/mchirico/gun-vs-vehicle-deaths/notebook>

2) Guns vs Vehicle related Deaths- Checking if people are double counted

<https://www.kaggle.com/jeffmoser/gun-vs-vehicle-deaths-exploring-query?scriptVersionId=185024>

The data seems to contradict what you read in the media:

1. Gun related deaths outnumber vehicle deaths.
2. White race related gun deaths are the highest.
3. White race and gun deaths seem somewhat correlated with vehicle deaths.

This query sees how many people have been double counted in death incidents

Are people double counted?  
SELECT ea.DeathRecordId, COUNT(\*)  
FROM EntityAxisConditions ea  
JOIN Icd10Code ic ON ea.Icd10Code = ic.Code  
WHERE ic.Description LIKE '%discharge%' AND NOT ic.Description LIKE '%rethral%' AND NOT ic.Description LIKE '%firework%' AND NOT ic.Description LIKE '%Legal intervention involving firearm discharge%'  
GROUP BY ea.DeathRecordId  
HAVING COUNT(\*) > 1  
ORDER BY COUNT(\*) DESC, DeathRecordId  
LIMIT 100

3) Age and Manner of Death

This query sees for each age ( 1, 2, 3, 4, 5) in which manner do people die

select  
Age,  
sum(b.Description == 'Not specified') as Not\_specified,  
sum(b.Description == 'Accident') as Accident,  
sum(b.Description == 'Suicide') as Suicide,  
sum(b.Description == 'Homicide') as Homicide,  
sum(b.Description == 'Pending investigation') as Pending\_investigation,  
sum(b.Description == 'Could not determine') as Could\_not\_determine,  
sum(b.Description == 'Self-Inflicted') as Self\_Inflicted,  
sum(b.Description == 'Natural') as Natural  
from DeathRecords a, MannerOfDeath b  
where a.MannerOfDeath = b.code  
and a.AgeType = 1  
group by a.Age  
order by age;

4) [Alzheimer's Deaths by Age Group](https://www.kaggle.com/jeffmoser/alzheimer-s-deaths-by-age-group)

https://www.kaggle.com/jeffmoser/alzheimer-s-deaths-by-age-group

SELECT ar.Description AS Age, COUNT(\*) AS TotalDeaths  
FROM DeathRecords dr  
JOIN AgeRecode12 ar ON ar.Code = dr.AgeRecode12  
JOIN Icd10Code ic ON ic.Code = dr.Icd10Code  
WHERE ic.Description LIKE '%Alzheimer%'  
GROUP BY ar.Description  
ORDER BY ar.Code

[How does education affect lifespan?](https://www.kaggle.com/jeffmoser/how-does-education-affect-lifespan)

SELECT er.Description AS Education, AVG(Age) AS AverageLifeExpectancy, COUNT(\*) AS TotalDeaths  
FROM DeathRecords dr  
JOIN Education2003Revision er ON er.Code = dr.Education2003Revision  
WHERE Education2003Revision IS NOT NULL AND AgeType = 1 AND Age > 35 AND AGE < 200  
GROUP BY dr.Education2003Revision

**Burial or cremation?**

**Code is in python**

<https://www.kaggle.com/igorsamokhin/burial-or-cremation>

**If interested, here’s a link on how to install spark on windows:** <https://medium.com/@GalarnykMichael/install-spark-on-windows-pyspark-4498a5d8d66c>

Python code:

1. US cause of death

<https://www.kaggle.com/jrskane/us-causes-of-death-analysis>

2) Men vs Women

[**https://www.kaggle.com/pablocastilla/men-vs-women**](https://www.kaggle.com/pablocastilla/men-vs-women)

**Dataset documentation:**

[**https://www.cdc.gov/nchs/data/dvs/Record\_Layout\_2014.pdf**](https://www.cdc.gov/nchs/data/dvs/Record_Layout_2014.pdf)