

Project Team Abele

Proposal

```
library(tidyverse)
```

Data 1

Introduction and data

- Identify the source of the data.
 - The source of the data is from the CORGIS Dataset Project.
- State when and how it was originally collected (by the original data curator, not necessarily how you found the data).
 - The data was originally collected by Dennis Kafura on June 27th, 2019 from patient records of various hospitals throughout the United States.
- Write a brief description of the observations.
 - The information provides rates of cancer deaths in each state. These rates are organized based on age, sex, and race. The rates also are categorized for three specific types of cancer: breast cancer, lung cancer, and colorectal cancer.

Research question

- A well formulated research question. (You may include more than one research question if you want to receive feedback on different ideas for your project. However, one per data set is required.)
 - What is the relationship between race and mortality rate and cancer types?

- A description of the research topic along with a concise statement of your hypotheses on this topic.
 - Hypotheses:
 - * Different racial groups have different mortality rates for the same type of cancer.
 - * The impact of race and cancer type on mortality rates is influenced by other factors such as age and gender.
 - Research topic:
 - * This research topic aims to explore the relationship between race, cancer type, and mortality among cancer patients. The first hypothesis suggests that mortality rates for the same type of cancer vary across different racial groups. For example, Black and Hispanic cancer patients may appear to have higher mortality rates for a certain type of cancer compared to White patients. The second hypothesis suggests that there are other factors such as age and gender that can impact the relationship between race, cancer type, and mortality rates. This research could potentially provide insights into how to improve cancer screening, diagnosis, and treatment strategies for different racial groups and different types of cancer.
- Identify the types of variables in your research question. Categorical? Quantitative?
 - The types of variables in our research question are categorical.

Literature

- Find one published credible article on the topic you are interested in researching.
 - Saldana-Ruiz, Nallely, et al. “Fundamental causes of colorectal cancer mortality in the United States: understanding the importance of socioeconomic status in creating inequality in mortality.” *American journal of public health* 103.1 (2013): 99-104.
- Provide a one paragraph summary about the article.
 - The goal of this study is to determine the creation of health disparities in colorectal cancer mortality in the United States from 1968 to 2005. Negative binomial regression was used to analyze trends based on county for colorectal cancer mortality rates, adjusted for gender, race, and age for individuals above 35 years old. The results reveal that prior to 1980, there was a stable gradient in colorectal cancer mortality, with higher SES counties at greater risk. However, beginning in 1980, this gradient began to narrow and then reversed as people living in higher SES counties

experiences greater reductions in colorectal cancer mortality rates than those in lower SES countries. This supports the fundamental cause hypothesis, suggesting that social and economic resources are important factors in influencing mortality rates once knowledge about prevention and treatment of colorectal cancer became available.

- In 1-2 sentences, explain how your research question builds on / is different than the article you have cited.
 - This research question builds on the research given because it also explores the relationship between other cancer types– lung cancer and breast cancer– as well as colorectal cancer.

Glimpse of data

```
cancer <- read_csv("data/cancer.csv")
```

```
Rows: 51 Columns: 75
```

```
-- Column specification -----
```

```
Delimiter: ","
```

```
chr (1): State
```

```
dbl (74): Total.Rate, Total.Number, Total.Population, Rates.Age.< 18, Rates....
```

```
i Use `spec()` to retrieve the full column specification for this data.
```

```
i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
glimpse(cancer)
```

```
Rows: 51
```

```
Columns: 75
```

```
$ State           <chr> "Alabama", "Alaska", "A~
$ Total.Rate      <dbl> 214.2, 128.1, 165.6, 22~
$ Total.Number    <dbl> 71529, 6361, 74286, 456~
$ Total.Population <dbl> 33387205, 4966180, 4484~
$ `Rates.Age.< 18` <dbl> 2.0, 1.7, 2.5, 2.3, 2.6~
$ `Rates.Age.18-45` <dbl> 18.5, 11.8, 13.6, 17.6,~
$ `Rates.Age.45-64` <dbl> 244.7, 170.9, 173.6, 25~
$ `Rates.Age.> 64`  <dbl> 1017.8, 965.2, 840.2, 1~
$ `Rates.Age and Sex.< 18.Female` <dbl> 2.0, 0.0, 2.6, 2.6, 2.4~
$ `Rates.Age and Sex.< 18.Male`  <dbl> 2.1, 0.0, 2.5, 2.0, 2.8~
```

\$ `Rates.Age and Sex.18 - 45.Female`	<dbl> 20.1, 13.9, 15.2, 19.5,~
\$ `Rates.Age and Sex.18 - 45.Male`	<dbl> 16.8, 10.0, 12.2, 15.8,~
\$ `Rates.Age and Sex.45 - 64.Female`	<dbl> 201.0, 157.6, 156.5, 21~
\$ `Rates.Age and Sex.45 - 64.Male`	<dbl> 291.5, 183.0, 191.8, 29~
\$ `Rates.Age and Sex.> 64.Female`	<dbl> 803.6, 849.6, 706.1, 83~
\$ `Rates.Age and Sex.> 64.Male`	<dbl> 1308.6, 1086.4, 1000.2,~
\$ Rates.Race.White	<dbl> 186.1, 168.2, 153.5, 19~
\$ `Rates.Race.White non-Hispanic`	<dbl> 187.5, 170.2, 156.4, 19~
\$ Rates.Race.Black	<dbl> 216.1, 183.7, 166.8, 22~
\$ Rates.Race.Asian	<dbl> 81.3, 118.7, 93.0, 115.~
\$ Rates.Race.Indigenous	<dbl> 69.9, 247.2, 116.2, 59.~
\$ `Rates.Race and Sex.Female.White`	<dbl> 149.2, 145.5, 130.8, 15~
\$ `Rates.Race and Sex.Female.White non-Hispanic`	<dbl> 150.2, 147.8, 134.0, 15~
\$ `Rates.Race and Sex.Female.Black`	<dbl> 167.2, 141.6, 146.5, 18~
\$ `Rates.Race and Sex.Female.Black non-Hispanic`	<dbl> 167.9, 148.8, 154.3, 18~
\$ `Rates.Race and Sex.Female.Asian`	<dbl> 84.9, 105.2, 80.6, 91.7~
\$ `Rates.Race and Sex.Female.Indigenous`	<dbl> 53.8, 219.4, 100.5, 44.~
\$ `Rates.Race and Sex.Male.White`	<dbl> 237.1, 195.6, 182.2, 24~
\$ `Rates.Race and Sex.Male.White non-Hispanic`	<dbl> 239.2, 197.3, 184.7, 24~
\$ `Rates.Race and Sex.Male.Black`	<dbl> 297.9, 240.8, 195.3, 30~
\$ `Rates.Race and Sex.Male.Black non-Hispanic`	<dbl> 299.2, 246.6, 205.0, 30~
\$ `Rates.Race and Sex.Male.Asian`	<dbl> 75.7, 142.8, 112.5, 157~
\$ `Rates.Race and Sex.Male.Indigenous`	<dbl> 88.3, 284.9, 139.5, 74.~
\$ Rates.Race.Hispanic	<dbl> 66.5, 88.8, 128.5, 81.2~
\$ `Rates.Race and Sex.Female.Hispanic`	<dbl> 58.3, 73.1, 106.5, 74.6~
\$ `Rates.Race and Sex.Male.Hispanic`	<dbl> 77.1, 104.0, 158.4, 90.~
\$ Types.Breast.Total	<dbl> 27.4, 17.8, 23.3, 27.9,~
\$ `Types.Breast.Age.18 - 44`	<dbl> 5.1, 2.9, 3.8, 5.0, 4.0~
\$ `Types.Breast.Age.45 - 64`	<dbl> 39.8, 27.8, 33.2, 38.4,~
\$ `Types.Breast.Age.> 64`	<dbl> 95.7, 108.0, 90.5, 100.~
\$ Types.Breast.Race.White	<dbl> 20.5, 21.3, 20.4, 21.7,~
\$ `Types.Breast.Race.White non-Hispanic`	<dbl> 20.6, 21.8, 20.9, 22.0,~
\$ Types.Breast.Race.Black	<dbl> 30.3, 0.0, 26.3, 31.1, ~
\$ `Types.Breast.Race.Black non-Hispanic`	<dbl> 30.4, 0.0, 28.0, 31.0, ~
\$ Types.Breast.Race.Asian	<dbl> 0.0, 12.2, 10.7, 0.0, 1~
\$ Types.Breast.Race.Indigenous	<dbl> 0.0, 25.5, 11.4, 0.0, 7~
\$ Types.Breast.Race.Hispanic	<dbl> 0.0, 0.0, 16.2, 10.9, 1~
\$ Types.Colorectal.Total	<dbl> 19.4, 11.9, 14.9, 21.2,~
\$ `Types.Colorectal.Age and Sex.Female.18 - 44`	<dbl> 1.6, 0.0, 1.3, 1.9, 1.2~
\$ `Types.Colorectal.Age and Sex.Male.18 - 44`	<dbl> 2.3, 0.0, 1.4, 2.2, 1.4~
\$ `Types.Colorectal.Age and Sex.Female.45 - 64`	<dbl> 18.0, 15.3, 12.6, 19.7,~
\$ `Types.Colorectal.Age and Sex.Male.45 - 64`	<dbl> 28.7, 17.6, 19.2, 28.6,~
\$ `Types.Colorectal.Age and Sex.Female.> 64`	<dbl> 78.4, 71.7, 67.6, 85.8,~

```

$ `Types.ColoRectal.Age and Sex.Male.> 64`      <dbl> 106.0, 102.3, 85.2, 114~
$ Types.ColoRectal.Race.White                   <dbl> 15.9, 13.6, 13.8, 17.7,~
$ `Types.ColoRectal.Race.White non-Hispanic`    <dbl> 16.0, 13.8, 13.9, 17.9,~
$ Types.ColoRectal.Race.Black                   <dbl> 24.4, 0.0, 18.7, 26.3, ~
$ `Types.ColoRectal.Race.Black non-Hispanic`    <dbl> 24.5, 0.0, 19.7, 26.4, ~
$ Types.ColoRectal.Race.Asian                   <dbl> 0.0, 12.5, 10.6, 0.0, 1~
$ Types.ColoRectal.Race.Indigenous              <dbl> 0.0, 34.7, 10.1, 0.0, 7~
$ Types.ColoRectal.Race.Hispanic                <dbl> 5.7, 0.0, 13.1, 8.1, 11~
$ Types.Lung.Total                             <dbl> 66.4, 36.6, 42.3, 73.3,~
$ `Types.Lung.Age and Sex.Female.18 - 44`      <dbl> 1.7, 0.0, 1.1, 1.9, 0.8~
$ `Types.Lung.Age and Sex.Male.18 - 44`        <dbl> 2.4, 0.0, 0.8, 1.8, 0.8~
$ `Types.Lung.Age and Sex.Female.45 - 64`      <dbl> 54.8, 39.7, 33.2, 61.4,~
$ `Types.Lung.Age and Sex.Male.45 - 64`        <dbl> 102.9, 50.3, 47.0, 106.~
$ `Types.Lung.Age and Sex.Female.> 64`         <dbl> 221.7, 268.3, 191.9, 24~
$ `Types.Lung.Age and Sex.Male.> 64`           <dbl> 457.4, 335.0, 275.8, 48~
$ Types.Lung.Race.White                        <dbl> 59.9, 48.7, 39.5, 63.4,~
$ `Types.Lung.Race.White non-Hispanic`         <dbl> 60.4, 49.5, 42.2, 64.2,~
$ Types.Lung.Race.Black                       <dbl> 52.6, 45.6, 38.2, 62.9,~
$ `Types.Lung.Race.Black non-Hispanic`         <dbl> 52.8, 47.9, 40.4, 63.0,~
$ Types.Lung.Race.Asian                       <dbl> 23.0, 33.0, 21.3, 18.1,~
$ Types.Lung.Race.Indigenous                  <dbl> 22.9, 74.4, 11.1, 16.2,~
$ Types.Lung.Race.Hispanic                    <dbl> 14.8, 0.0, 21.6, 14.6, ~

```

Data 2

Introduction and data

- Identify the source of the data.
 - This data is sourced from the CORGIS Dataset Project and was uploaded by Dennis Kafura, Joung Min Choi, and Bo Guan.
- State when and how it was originally collected (by the original data curator, not necessarily how you found the data).
 - This dataset was originally collected by the Washington Post over the span of 2015-2022. They assembled the dataset by manually culling local news reports, amassing information from social media and law enforcement websites, and monitoring independent databases such as Fatal Encounters.
- Write a brief description of the observations.

- The dataset contains observations about instances of police shootings in the US. The variables are mostly focused on the details of the situation and people involved in the shooting, such as the date and location of the shooting and whether the victim was armed or experiencing a mental health crisis.

Research question

- A well formulated research question. (You may include more than one research question if you want to receive feedback on different ideas for your project. However, one per data set is required.)
 - Is there a relationship between the race of the person shot and existence of a perceived threat and presence of mental illness during police shootings in the United States?
- A description of the research topic along with a concise statement of your hypotheses on this topic.
 - A recent uptick in police brutality and racially motivated police action has sparked outrage nationwide. In fact, the creation of this data set was prompted by the widespread calls for increased police accountability and the current gap in tracking of the situational and interpersonal, specifically racial, details surrounding police shootings. Thus, our research seeks to explore the relationship between the race of the victim and the police's level of perceived threat and perceived presence of a mental illness.
 - We hypothesize that there is a statistically significant relationship between the race of the victim and the perceived threat level and perceived mental illness. Specifically, we hypothesize that minority races, specifically African American, will correlate with a higher level of perceived threat and perceived presence of a mental illness.
- Identify the types of variables in your research question. Categorical? Quantitative?
 - The variables Person.Race, Factors.Mental-Illness, and Factors.Armed are all categorical.

Literature

- Find one published credible article on the topic you are interested in researching.

- Examining Individual and Aggregate Correlates of Police Killings of People with Mental Illness: A Special Gaze at Race and Ethnicity. Prince, K. J., & Sun, I. Y. (2023). Examining Individual and Aggregate Correlates of Police Killings of People with Mental Illness: A Special Gaze at Race and Ethnicity. *Homicide Studies*, 27(1), 77–96. <https://doi.org/10.1177/10887679221119397>
- Provide a one paragraph summary about the article.
 - This article studies the influences of race on the existence of resistant behavior in instances of deadly police contact in people who possess mental illnesses. There are many studies focusing on civilian-police conflict, but few focus on instances in which the civilian has a mental illness and the effects of that on the encounter. Overall, they found that race was not a significant or consistent predictor of resistant behavior among civilians. They found that current literature examining the influence of mental illness on police and civilian interactions sheds light on potential explanations for resistance but still leaves many gaps to fill.
- In 1-2 sentences, explain how your research question builds on / is different than the article you have cited.
 - Our research question builds on the data explored in the article because it includes instances of civilian victims who both possess and don't possess mental illnesses. It also differs from the article in that it does not discuss resistant behavior as a component of deadly police interactions.

Glimpse of data

```
police_shootings <- read_csv("data/police_shootings.csv")
```

```
Rows: 6569 Columns: 16
```

```
-- Column specification -----
```

```
Delimiter: ","
```

```
chr  (9): Person.Name, Person.Gender, Person.Race, Incident.Location.City, I...
```

```
dbl  (4): Person.Age, Incident.Date.Month, Incident.Date.Day, Incident.Date....
```

```
lgl  (2): Factors.Mental-Illness, Shooting.Body-Camera
```

```
date (1): Incident.Date.Full
```

```
i Use `spec()` to retrieve the full column specification for this data.
```

```
i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
glimpse(police_shootings)
```

```

Rows: 6,569
Columns: 16
$ Person.Name          <chr> "Tim Elliot", "Lewis Lee Lembke", "John Paul ~
$ Person.Age           <dbl> 53, 47, 23, 32, 39, 18, 22, 35, 34, 47, 25, 3~
$ Person.Gender        <chr> "Male", "Male", "Male", "Male", "Male", "Male~
$ Person.Race          <chr> "Asian", "White", "Hispanic", "White", "Hispa~
$ Incident.Date.Month  <dbl> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, ~
$ Incident.Date.Day    <dbl> 2, 2, 3, 4, 4, 4, 5, 6, 6, 6, 6, 7, 7, 7, ~
$ Incident.Date.Year   <dbl> 2015, 2015, 2015, 2015, 2015, 2015, 2015, 201~
$ Incident.Date.Full   <date> 2015-01-02, 2015-01-02, 2015-01-03, 2015-01-~
$ Incident.Location.City <chr> "Shelton", "Aloha", "Wichita", "San Francisco~
$ Incident.Location.State <chr> "WA", "OR", "KS", "CA", "CO", "OK", "AZ", "KS~
$ Factors.Armed        <chr> "gun", "gun", "unarmed", "toy weapon", "nail ~
$ `Factors.Mental-Illness` <lgl> TRUE, TRUE, TRUE, TRUE, TRUE, TRUE, TRUE, TRU~
$ `Factors.Threat-Level` <chr> "attack", "attack", "other", "attack", "attac~
$ Factors.Fleeing      <chr> "Not fleeing", "Not fleeing", "Not fleeing", ~
$ Shooting.Manner      <chr> "shot", "shot", "shot and Tasered", "shot", "~
$ `Shooting.Body-Camera` <lgl> TRUE, TRUE, TRUE, TRUE, TRUE, TRUE, TRUE, TRU~

```

Data 3

Introduction and data

- Identify the source of the data.
 - The Opioids dataset is taken from the CORGIS Dataset Project.
- State when and how it was originally collected (by the original data curator, not necessarily how you found the data).
 - It was collected by the National Institute on Drug Abuse through resources like emergency room and rehabilitation center data, and records data about opioid abuse and opioid related deaths between the years 1999 and 2019.
- Write a brief description of the observations.
 - The observations are broken down to examine the prevalence of opioid abuse into a number of factors, such as the type of opioid abused, and race and sex of the abuser.

Research question

- A well formulated research question. (You may include more than one research question if you want to receive feedback on different ideas for your project. However, one per data set is required.)
 - How does race affect the rate of opioid abuse death in the US?
- A description of the research topic along with a concise statement of your hypotheses on this topic.
 - In recent years, the United States has been afflicted by an opioid crisis which has seen opioid overdose deaths increase exponentially from 21,089 in 2010 to more than 106,000 in 2021. It has become one of the biggest killers in the U.S., especially of men. In light of this, we want to investigate whether race has an impact on the rate of opioid abuse death. We hypothesize that opioid abuse death will be most prevalent among White, Black and Hispanic people, and less prevalent among American Indian or Alaskan Native and Asian or Pacific Islander people.
- Identify the types of variables in your research question. Categorical? Quantitative?
 - Rate.Opioid.Race.Hispanic
 - Rate.Opioid.Race.Asian or Pacific Islander
 - Rate.Opioid.Black
 - Rate.Opioid.White
 - Rate.Opioid.American Indian or Alaska Native
 - Rate.Opioid.Any.Total
 - These variables are all quantitative. All but the last one records the rate of overdose deaths among the specified racial group due to any opioid per 100,000 people. The final variable records the rate of overdose deaths due to any opioid per 100,000 people.

Literature

- Find one published credible article on the topic you are interested in researching.
 - We are going to refer to a 2008 article “Trends in Opioid Prescribing by Race/Ethnicity for Patients Seeking Care in US Emergency Departments” to complement our research.
- Provide a one paragraph summary about the article.

- This article looks at how race affects the likelihood of cocaine and psychostimulants overdose deaths. It found that death rates were higher for Black people and American Indian/Alaskan Native people. It also found that cocaine and opioid overdoses increased among Hispanic, White and Black people, as did psychostimulants overdoses.
- In 1-2 sentences, explain how your research question builds on / is different than the article you have cited.
 - Our research question does not focus on cocaine and psychostimulant overdoses but instead opioid overdoses. It also considers data over a longer period of time, from 1999-2019 as opposed to 2004-2019. In this way, we hope to use this article to compare against opioid trends.

Glimpse of data

```
opioids <- read_csv("data/opioids.csv")
```

```
Rows: 21 Columns: 49
```

```
-- Column specification -----
```

```
Delimiter: ","
```

```
dbl (49): Year, Number.All, Number.Opioid.Any, Number.Opioid.Prescription, N...
```

```
i Use `spec()` to retrieve the full column specification for this data.
```

```
i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
glimpse(opioids)
```

```
Rows: 21
```

```
Columns: 49
```

\$ Year	<dbl> 1999, ~
\$ Number.All	<dbl> 16849~
\$ Number.Opioid.Any	<dbl> 8050, ~
\$ Number.Opioid.Prescription	<dbl> 3442, ~
\$ Number.Opioid.Synthetic	<dbl> 730, ~
\$ Number.Opioid.Heroin	<dbl> 1960, ~
\$ Number.Opioid.Cocaine	<dbl> 3822, ~
\$ Rate.All.Total	<dbl> 6.1, ~
\$ Rate.All.Sex.Female	<dbl> 3.9, ~
\$ Rate.All.Sex.Male	<dbl> 8.2, ~

\$ Rate.All.Race.White	<dbl> 6.2, ~
\$ Rate.All.Race.Black	<dbl> 7.5, ~
\$ `Rate.All.Race.Asian or Pacific Islander`	<dbl> 1.2, ~
\$ Rate.All.Race.Hispanic	<dbl> 5.4, ~
\$ `Rate.All.Race.American Indian or Alaska Native`	<dbl> 6.0, ~
\$ Rate.Opioid.Any.Total	<dbl> 2.9, ~
\$ Rate.Opioid.Any.Sex.Female	<dbl> 1.4, ~
\$ Rate.Opioid.Any.Sex.Male	<dbl> 4.3, ~
\$ Rate.Opioid.Any.Race.White	<dbl> 2.8, ~
\$ Rate.Opioid.Any.Race.Black	<dbl> 3.5, ~
\$ `Rate.Opioid.Any.Race.Asian or Pacific Islander`	<dbl> 0.3, ~
\$ Rate.Opioid.Any.Race.Hispanic	<dbl> 3.5, ~
\$ `Rate.Opioid.Any.Race.American Indian or Alaska Native`	<dbl> 2.9, ~
\$ Rate.Opioid.Prescription.Total	<dbl> 1.2, ~
\$ Rate.Opioid.Prescription.Sex.Female	<dbl> 0.7, ~
\$ Rate.Opioid.Prescription.Sex.Male	<dbl> 1.7, ~
\$ Rate.Opioid.Prescription.Race.White	<dbl> 1.3, ~
\$ Rate.Opioid.Prescription.Race.Black	<dbl> 0.8, ~
\$ Rate.Opioid.Prescription.Race.Hispanic	<dbl> 1.6, ~
\$ `Rate.Opioid.Prescription.Race.American Indian or Alaska Native`	<dbl> 1.3, ~
\$ Rate.Opioid.Synthetic.Total	<dbl> 0.3, ~
\$ Rate.Opioid.Synthetic.Sex.Female	<dbl> 0.2, ~
\$ Rate.Opioid.Synthetic.Sex.Male	<dbl> 0.3, ~
\$ Rate.Opioid.Synthetic.Race.White	<dbl> 0.3, ~
\$ Rate.Opioid.Synthetic.Race.Black	<dbl> 0.1, ~
\$ Rate.Opioid.Synthetic.Race.Hispanic	<dbl> 0.1, ~
\$ Rate.Opioid.Heroin.Total	<dbl> 0.7, ~
\$ Rate.Opioid.Heroin.Sex.Female	<dbl> 0.2, ~
\$ Rate.Opioid.Heroin.Sex.Male	<dbl> 1.2, ~
\$ Rate.Opioid.Heroin.Race.White	<dbl> 0.7, ~
\$ Rate.Opioid.Heroin.Race.Black	<dbl> 0.8, ~
\$ Rate.Opioid.Heroin.Race.Hispanic	<dbl> 1.1, ~
\$ Rate.Opioid.Cocaine.Total	<dbl> 1.4, ~
\$ Rate.Opioid.Cocaine.Sex.Female	<dbl> 0.6, ~
\$ Rate.Opioid.Cocaine.Sex.Male	<dbl> 2.1, ~
\$ Rate.Opioid.Cocaine.Race.White	<dbl> 1.0, ~
\$ Rate.Opioid.Cocaine.Race.Black	<dbl> 3.7, ~
\$ Rate.Opioid.Cocaine.Race.Hispanic	<dbl> 1.7, ~
\$ `Rate.Opioid.Cocaine.Race.American Indian or Alaska Native`	<dbl> 0.9, ~