Manuscript Template for a Data Analysis Project

Exploring Factors influencing HIV-related Mortality Rates in the Georgia – 2018 - 2020

Hayley Hemme

2/9/23

The structure below is one possible setup for a data analysis project (including the course project). For a manuscript, adjust as needed. You don’t need to have exactly these sections, but the content covering those sections should be addressed.

This uses MS Word as output format. [See here](https://quarto.org/docs/output-formats/ms-word.html) for more information. You can switch to other formats, like html or pdf. See [the Quarto documentation](https://quarto.org/) for other formats.

Warning: package 'here' was built under R version 4.2.2

Warning: package 'knitr' was built under R version 4.2.2

Warning: package 'tidyverse' was built under R version 4.2.2

Warning: package 'ggplot2' was built under R version 4.2.2

Warning: package 'tidyr' was built under R version 4.2.2

Warning: package 'readr' was built under R version 4.2.2

Warning: package 'purrr' was built under R version 4.2.2

Warning: package 'dplyr' was built under R version 4.2.2

Warning: package 'stringr' was built under R version 4.2.2

Warning: package 'forcats' was built under R version 4.2.2

Warning: package 'readxl' was built under R version 4.2.2

Warning: package 'lubridate' was built under R version 4.2.2

[1] "C:/Users/Hayley/Desktop/MADA2023/HAYLEYHEMME-MADA-project"

# 1. Summary/Abstract

*Write a summary of your project.*

# 2. Introduction

## 2.1 General Background Information

*Provide enough background on your topic that others can understand the why and how of your analysis*

## 2.2 Description of data and data source

The data used in this project was obtained from multiple publicly available sources. Data pertaining to counts of HIV-related deaths in for Hispanics, non-Hispanic Whites, and non-Hispanic Blacks and populations for these groups was obtained from [OASIS](https://oasis.state.ga.us/), the Georgia Department of Public Health’s online data repository. This analyse will currently focus on the year 2018-2020, but may be expanded to 2016-2020 moving forward. The data on these years will be obtained and processed identically to ways later described. Date on new HIV diagnoses and SDOH, particularly related to HIV were obtained from [AIDSVu](https://aidsvu.org/resources/#/datasets). Data on poverty rates was obtained from [Georgia Data](https://georgiadata.org/), a data source provided by the the Carl Vinson Institute of Government at the University of Georgia. All data described was aggregated at the county level.

## 2.3 Questions/Hypotheses to be addressed

Question: Can HIV death rates be predicted by social determinants of health like poverty and/ or by HIV related outcomes (new diagnoses, prevalence PrEP use, etc?)

H0 : There is no correlation between HIV-related deaths and SDOH or HIV-related outcomes. Ha : There is a correlation between HIV-related deaths and SDOH or HIV-related outcomes.

To cite other work (important everywhere, but likely happens first in introduction), make sure your references are in the bibtex file specified in the YAML header above (here dataanalysis\_template\_references.bib) and have the right bibtex key. Then you can include like this:

Examples of reproducible research projects can for instance be found in (McKay, Ebell, Billings, et al., 2020; McKay, Ebell, Dale, Shen, & Handel, 2020)

# 3. Methods

*Describe your methods. That should describe the data, the cleaning processes, and the analysis approaches. You might want to provide a shorter description here and all the details in the supplement.*

## 3.1 Data aquisition

*As applicable, explain where and how you got the data. If you directly import the data from an online source, you can combine this section with the next.*

## 3.2 Data import and cleaning

*Write code that reads in the file and cleans it so it’s ready for analysis. Since this will be fairly long code for most datasets, it might be a good idea to have it in one or several R scripts. If that is the case, explain here briefly what kind of cleaning/processing you do, and provide more details and well documented code somewhere (e.g. as supplement in a paper). All materials, including files that contain code, should be commented well so everyone can follow along.* Loading the datasets from OASIS for deaths and population (Hispanic, NH White, NH Black) for persons ages 13-59 from 2018 to 2020.

his\_deaths <- read\_xlsx(here("data/raw\_data/hivdeaths\_his\_1820.xlsx"), skip = 1)  
nhw\_deaths <- read\_xlsx(here("data/raw\_data/hivdeaths\_nhw\_1820.xlsx"), skip = 1)  
nhb\_deaths <- read\_xlsx(here("data/raw\_data/hivdeaths\_nhb\_1820.xlsx"), skip = 1)  
  
his\_pop <- read\_xlsx(here("data/raw\_data/pop\_his\_1820.xlsx"), skip = 1)  
nhw\_pop <- read\_xlsx(here("data/raw\_data/pop\_nhw\_1820.xlsx"), skip = 1)  
nhb\_pop <- read\_xlsx(here("data/raw\_data/pop\_nhb\_1820.xlsx"), skip = 1)

Loading dataset from AIDSVu for social determinants of health (2020) and new HIV diagnoses from 2018 to 2020.

sdoh <- read\_xlsx(here("data/raw\_data/AIDSVu\_County\_SDOH\_2020.xlsx"), skip = 3)  
sdoh <- sdoh %>% filter(State == "Georgia")  
  
dx18 <- read\_xlsx(here("data/raw\_data/AIDSVu\_County\_NewDX\_2018.xlsx"), skip = 3)  
dx18 <- dx18 %>% filter(State == "Georgia") %>%   
 select(!contains(c("Rate Stability", "Percent", "Case")))  
  
dx19 <- read\_xlsx(here("data/raw\_data/AIDSVu\_County\_NewDX\_2019.xlsx"), skip = 3)  
dx19 <- dx19 %>% filter(State == "Georgia")%>%   
 select(!contains(c("Rate Stability", "Percent", "Case")))  
  
dx20 <- read\_xlsx(here("data/raw\_data/AIDSVu\_County\_NewDX\_2020.xlsx"), skip = 3)  
dx20 <- dx20 %>% filter(State == "Georgia") %>%   
 select(!contains(c("Rate Stability", "Percent", "Case")))

Let’s load data from Georgia Data on Poverty rates in Georgia by race/ ethnicity for the period 2016-2020.

pov <- read\_xlsx(here("data/raw\_data/economics\_22.xlsx"), sheet = "poverty")   
  
glimpse(pov)

Rows: 168  
Columns: 23  
$ County <chr> …  
$ `2020 Total Persons in Poverty` <dbl> …  
$ `2020 Total Persons in Poverty, Percent` <dbl> …  
$ `2020 Rank of Total Persons in Poverty, Percent` <chr> …  
$ `2020 Children Under Age 18 in Poverty` <dbl> …  
$ `2020 Children Under Age 18 in Poverty, Percent` <dbl> …  
$ `2020 Related Children in Families, Age 5-17, in Poverty` <dbl> …  
$ `2020 Related Children in Families, Age 5-17, in Poverty, Percent` <dbl> …  
$ `2016-2020 Persons Below Poverty Level, Percent` <dbl> …  
$ `2016-2020 Persons Below Poverty Level, White, Percent` <dbl> …  
$ `2016-2020 Persons Below Poverty Level, Black, Percent` <dbl> …  
$ `2016-2020 Persons Below Poverty Level, Hispanic, Percent` <chr> …  
$ `2016-2020 Persons Age 65 and Over in Poverty, Percent` <dbl> …  
$ `2016-2020 Persons Less than High School Graduates in Poverty, Percent` <dbl> …  
$ `2016-2020 Persons with Bachelor's Degree or Higher in Poverty, Percent` <dbl> …  
$ `2016-2020 Families Below Poverty Level, Percent` <dbl> …  
$ `2016-2020 Families Below Poverty Level with a White Householder, Percent` <dbl> …  
$ `2016-2020 Families Below Poverty Level with a Black Householder, Percent` <chr> …  
$ `2016-2020 Families Below Poverty Level with a Hispanic Householder, Percent` <chr> …  
$ `2016-2020 Families Below Poverty Level, Female Head of Household, No Husband Present, Percent` <dbl> …  
$ `2016-2020 Poverty Level with a White Householder, Female Head of Household, No Husband Present, Percent` <dbl> …  
$ `2016-2020 Poverty Level with a Black Householder, Female Head of Household, No Husband Present, Percent` <chr> …  
$ `2016-2020 Poverty Level with an Hispanic Householder, Female Head of Household, No Husband Present, Percent` <chr> …

We’ll just be using the data from the “2016-2020 Persons Below Poverty Level” by race/ethnicity. We’ll also change the text formatting of County to match the other data frames.

pov <- pov %>% select(1, starts\_with("2016-2020 Persons Below Poverty Level"))  
  
pov$County <- str\_to\_title(pov$County)

Let’s do some cleaning on the OASIS data. We’ll remove extra columns and rows, as well as rows with “Selected ages/sexes total”, pivot Year and Deaths into long format, and create the new variable for race/ethnicity.

his\_deaths <- his\_deaths[c(-1,-2387:-2414),-7]  
his\_deaths <- his\_deaths %>% filter(!grepl("Selected", Age)) %>% filter(!grepl("Selected", Sex)) %>%  
 pivot\_longer(c(4:6), names\_to = "Year", values\_to = "HIV\_Deaths") %>%   
 mutate("Race\_Ethnicity" = "Hispanic")  
  
nhw\_deaths <- nhw\_deaths[c(-1,-2387:-2414),-7]  
nhw\_deaths <- nhw\_deaths %>% filter(!grepl("Selected", Age)) %>% filter(!grepl("Selected", Sex)) %>%  
 pivot\_longer(c(4:6), names\_to = "Year", values\_to = "HIV\_Deaths") %>%   
 mutate("Race\_Ethnicity" = "NH White")  
  
nhb\_deaths <- nhb\_deaths[c(-1,-2387:-2414),-7]  
nhb\_deaths <- nhb\_deaths %>% filter(!grepl("Selected", Age)) %>% filter(!grepl("Selected", Sex)) %>%  
 pivot\_longer(c(4:6), names\_to = "Year", values\_to = "HIV\_Deaths") %>%   
 mutate("Race\_Ethnicity" = "NH Black")

Next, we’ll merge the datasets into one dataframe, and remove the original datasets from the environment.

deaths <- full\_join(his\_deaths, nhw\_deaths)

Joining with `by = join\_by(Geography, Age, Sex, Year, HIV\_Deaths,  
Race\_Ethnicity)`

deaths <- full\_join(deaths, nhb\_deaths)

Joining with `by = join\_by(Geography, Age, Sex, Year, HIV\_Deaths,  
Race\_Ethnicity)`

rm(his\_deaths, nhw\_deaths, nhb\_deaths)

Next, let’s clean up the dataset for population using a similar method.

his\_pop <- his\_pop[c(-1,-2387:-2412),-7]  
his\_pop <- his\_pop %>% filter(!grepl("Selected", Age)) %>% filter(!grepl("Selected", Sex)) %>%   
 pivot\_longer(c(4:6), names\_to = "Year", values\_to = "Population") %>%   
 mutate("Race\_Ethnicity" = "Hispanic")   
  
nhw\_pop <- nhw\_pop[c(-1,-2387:-2412),-7]  
nhw\_pop <- nhw\_pop %>% filter(!grepl("Selected", Age)) %>% filter(!grepl("Selected", Sex)) %>%   
 pivot\_longer(c(4:6), names\_to = "Year", values\_to = "Population") %>%   
 mutate("Race\_Ethnicity" = "NH White")  
  
nhb\_pop <- nhb\_pop[c(-1,-2387:-2412),-7]  
nhb\_pop <- nhb\_pop %>% filter(!grepl("Selected", Age)) %>% filter(!grepl("Selected", Sex)) %>%   
 pivot\_longer(c(4:6), names\_to = "Year", values\_to = "Population") %>%   
 mutate("Race\_Ethnicity" = "NH Black")

Merging datasets and cleaning the original from the environment.

pop <- full\_join(his\_pop, nhw\_pop)

Joining with `by = join\_by(Geography, Age, Sex, Year, Population,  
Race\_Ethnicity)`

pop <- full\_join(pop, nhb\_pop)

Joining with `by = join\_by(Geography, Age, Sex, Year, Population,  
Race\_Ethnicity)`

rm(his\_pop, nhw\_pop, nhb\_pop)

Next, let’s merge the two OASIS dataframes.

df <- full\_join(deaths, pop, by = c("Geography", "Sex", "Age", "Year", "Race\_Ethnicity"))

Next, let’s create a new variable for Public Health District, change some variable to factors, and rename the Age variable and Geography variable.

df <- df %>% rename(County = Geography) %>%   
 mutate(District = case\_when(County %in% c("Bartow", "Catoosa", "Chattooga", "Dade", "Floyd", "Gordon", "Haralson", "Paulding", "Polk", "Walker") ~ "1-1",  
 County %in% c("Cherokee", "Fannin", "Gilmer", "Murray", "Pickens", "Whitfield") ~ "1-2",  
 County %in% c("Banks", "Dawson", "Forsyth", "Franklin", "Habersham", "Hall", "Hart", "Lumpkin", "Rabun", "Stephens", "Towns", "Union", "White") ~ "2",  
 County %in% c("Cobb", "Douglas") ~ "3-1",  
 County %in% c("Fulton") ~ "3-2",  
 County %in% c("Clayton") ~ "3-3",  
 County %in% c("Gwinnett", "Newton", "Rockdale") ~ "3-4",  
 County %in% c("DeKalb") ~ "3-5",  
 County %in% c("Butts", "Carroll", "Coweta", "Fayette", "Heard", "Henry", "Lamar", "Meriwether", "Pike", "Spalding", "Troup", "Upson") ~ "4",  
 County %in% c("Bleckley", "Dodge", "Johnson", "Laurens", "Montgomery", "Pulaski", "Telfair", "Treutlen", "Wheeler", "Wilcox") ~ "5-1",  
 County %in% c("Baldwin", "Bibb", "Crawford", "Hancock", "Houston", "Jasper", "Jones", "Monroe", "Peach", "Putnam", "Twiggs", "Washington", "Wilkinson") ~ "5-2",  
 County %in% c("Burke", "Columbia", "Emanuel", "Glascock", "Jefferson", "Jenkins", "Lincoln", "McDuffie", "Richmond", "Screven", "Taliaferro", "Warren", "Wilkes") ~ "6",  
 County %in% c("Chattahoochee", "Clay", "Crisp", "Dooly", "Harris", "Macon", "Muscogee", "Marion", "Quitman", "Randolph", "Schley", "Stewart", "Sumter", "Talbot", "Taylor", "Webster") ~ "7",  
 County %in% c("Ben Hill", "Berrien", "Brooks", "Cook", "Echols", "Irwin", "Lanier", "Lowndes", "Tift", "Turner") ~ "8-1",  
 County %in% c("Baker", "Calhoun", "Colquitt", "Dougherty", "Decatur", "Early", "Grady", "Lee", "Miller", "Mitchell", "Seminole", "Terrell", "Thomas", "Worth") ~ "8-2",  
 County %in% c("Bryan", "Camden", "Chatham", "Effingham", "Glynn", "Liberty", "Long", "McIntosh") ~ "9-1",  
 County %in% c("Appling", "Atkinson", "Bacon", "Brantley", "Bulloch", "Candler", "Charlton", "Clinch", "Coffee", "Evans", "Jeff Davis", "Pierce", "Tattnall", "Toombs", "Ware", "Wayne") ~ "9-2",  
 County %in% c("Barrow", "Clarke", "Elbert", "Greene", "Jackson", "Madison", "Morgan", "Oconee", "Oglethorpe", "Walton") ~ "10")) %>%   
 mutate(District = as.factor(District)) %>%   
 mutate(Sex = as.factor(Sex)) %>%   
 mutate(Age = as.factor(Age)) %>%   
 mutate(County = as.factor(County)) %>%   
 mutate(Race\_Ethnicity = as.factor(Race\_Ethnicity)) %>%   
 rename(Age\_Group = Age) %>%   
 mutate(Age\_Group = as.factor(Age\_Group))

Let’s clean up the Age\_Group variable by removing text

df$Age\_Group <- str\_replace(df$Age\_Group, " Adolescence", "")  
df$Age\_Group <- str\_replace(df$Age\_Group, " Early Adulthood", "")  
df$Age\_Group <- str\_replace(df$Age\_Group, " Young Adulthood", "")  
df$Age\_Group <- str\_replace(df$Age\_Group, " Middle Adulthood", "")

glimpse(df)

Rows: 11,448  
Columns: 8  
$ County <fct> Appling, Appling, Appling, Appling, Appling, Appling, A…  
$ Age\_Group <chr> "13-19", "13-19", "13-19", "13-19", "13-19", "13-19", "…  
$ Sex <fct> Male, Male, Male, Female, Female, Female, Male, Male, M…  
$ Year <chr> "2018", "2019", "2020", "2018", "2019", "2020", "2018",…  
$ HIV\_Deaths <chr> "0", "0", "0", "0", "0", "0", "0", "0", "2", "0", "0", …  
$ Race\_Ethnicity <fct> Hispanic, Hispanic, Hispanic, Hispanic, Hispanic, Hispa…  
$ Population <chr> "97", "96", "112", "119", "113", "119", "161", "154", "…  
$ District <fct> 9-2, 9-2, 9-2, 9-2, 9-2, 9-2, 9-2, 9-2, 9-2, 9-2, 9-2, …

Things are looking good so far!

Now let’s move onto the AIDSVu data. First, we’ll merge all new HIV diagnoses from 2018-2020 into one dataframe

dx <- full\_join(dx18, dx19)

Joining with `by = join\_by(`GEO ID`, Year, `State Abbreviation`, State, `County  
Name`, `New Diagnoses Rate`, `New Diagnoses Male Rate`, `New Diagnoses Female  
Rate`, `New Diagnoses Black Rate`, `New Diagnoses White Rate`, `New Diagnoses  
Hispanic Rate`, `New Diagnoses Asian Rate`, `New Diagnoses American  
Indian/Alaska Native Rate`, `New Diagnoses Multiracial Rate`, `New Diagnoses  
Native Hawaiian/Pacific Islander Rate`, `New Diagnoses Age 13-24 Rate`, `New  
Diagnoses Age 25-34 Rate`, `New Diagnoses Age 35-44 Rate`, `New Diagnoses Age  
45-54 Rate`, `New Diagnoses Age 55+ Rate`, `New Diagnoses MSM Rate`, `2013 NCHS  
Urbanicity Code`)`

dx <- full\_join(dx, dx20)

Joining with `by = join\_by(`GEO ID`, Year, `State Abbreviation`, State, `County  
Name`, `New Diagnoses Rate`, `New Diagnoses Male Rate`, `New Diagnoses Female  
Rate`, `New Diagnoses Black Rate`, `New Diagnoses White Rate`, `New Diagnoses  
Hispanic Rate`, `New Diagnoses Asian Rate`, `New Diagnoses American  
Indian/Alaska Native Rate`, `New Diagnoses Multiracial Rate`, `New Diagnoses  
Native Hawaiian/Pacific Islander Rate`, `New Diagnoses Age 13-24 Rate`, `New  
Diagnoses Age 25-34 Rate`, `New Diagnoses Age 35-44 Rate`, `New Diagnoses Age  
45-54 Rate`, `New Diagnoses Age 55+ Rate`, `New Diagnoses MSM Rate`, `2013 NCHS  
Urbanicity Code`)`

Let’s remove columns for race/ethnicities other than Hispanic, NH White, and NH Black, as well as some other variables we won’t be using in this analysis. We will also rename county and create the district variable.

dx <- dx %>% select(!contains(c("Asian", "American", "Multiracial", "Hawaiian", "MSM", "2013", "GEO ID", "State")))%>%   
 rename(County = "County Name")   
  
dx$Geography <- str\_remove(dx$County, " County")  
  
dx <- dx %>% mutate(District = case\_when(County %in% c("Bartow", "Catoosa", "Chattooga", "Dade", "Floyd", "Gordon", "Haralson", "Paulding", "Polk", "Walker") ~ "1-1",  
 County %in% c("Cherokee", "Fannin", "Gilmer", "Murray", "Pickens", "Whitfield") ~ "1-2",  
 County %in% c("Banks", "Dawson", "Forsyth", "Franklin", "Habersham", "Hall", "Hart", "Lumpkin", "Rabun", "Stephens", "Towns", "Union", "White") ~ "2",  
 County %in% c("Cobb", "Douglas") ~ "3-1",  
 County %in% c("Fulton") ~ "3-2",  
 County %in% c("Clayton") ~ "3-3",  
 County %in% c("Gwinnett", "Newton", "Rockdale") ~ "3-4",  
 County %in% c("DeKalb") ~ "3-5",  
 County %in% c("Butts", "Carroll", "Coweta", "Fayette", "Heard", "Henry", "Lamar", "Meriwether", "Pike", "Spalding", "Troup", "Upson") ~ "4",  
 County %in% c("Bleckley", "Dodge", "Johnson", "Laurens", "Montgomery", "Pulaski", "Telfair", "Treutlen", "Wheeler", "Wilcox") ~ "5-1",  
 County %in% c("Baldwin", "Bibb", "Crawford", "Hancock", "Houston", "Jasper", "Jones", "Monroe", "Peach", "Putnam", "Twiggs", "Washington", "Wilkinson") ~ "5-2",  
 County %in% c("Burke", "Columbia", "Emanuel", "Glascock", "Jefferson", "Jenkins", "Lincoln", "McDuffie", "Richmond", "Screven", "Taliaferro", "Warren", "Wilkes") ~ "6",  
 County %in% c("Chattahoochee", "Clay", "Crisp", "Dooly", "Harris", "Macon", "Muscogee", "Marion", "Quitman", "Randolph", "Schley", "Stewart", "Sumter", "Talbot", "Taylor", "Webster") ~ "7",  
 County %in% c("Ben Hill", "Berrien", "Brooks", "Cook", "Echols", "Irwin", "Lanier", "Lowndes", "Tift", "Turner") ~ "8-1",  
 County %in% c("Baker", "Calhoun", "Colquitt", "Dougherty", "Decatur", "Early", "Grady", "Lee", "Miller", "Mitchell", "Seminole", "Terrell", "Thomas", "Worth") ~ "8-2",  
 County %in% c("Bryan", "Camden", "Chatham", "Effingham", "Glynn", "Liberty", "Long", "McIntosh") ~ "9-1",  
 County %in% c("Appling", "Atkinson", "Bacon", "Brantley", "Bulloch", "Candler", "Charlton", "Clinch", "Coffee", "Evans", "Jeff Davis", "Pierce", "Tattnall", "Toombs", "Ware", "Wayne") ~ "9-2",  
 County %in% c("Barrow", "Clarke", "Elbert", "Greene", "Jackson", "Madison", "Morgan", "Oconee", "Oglethorpe", "Walton") ~ "10")) %>%   
 mutate(District = as.factor(District))

Things are looking pretty good so far, so let’s save these dataframes.

saveRDS(df, here("data/processed\_data/df.rds"))  
saveRDS(dx, here("data/processed\_data/dx.rds"))  
saveRDS(sdoh, here("data/processed\_data/sdoh.rds"))  
saveRDS(pov, here("data/processed\_data/pov.rds"))

Clear the environment and start fresh.

## 3.3 Statistical analysis

Explore central tendency across demographic groups. Fit a poisson regression model predicting death-to-case ratio from percentage with x SDOH variables and percentage of x demographic group.

# 4. Results

## 4.1 Exploratory/Descriptive analysis

*Use a combination of text/tables/figures to explore and describe your data. Show the most important descriptive results here. Additional ones should go in the supplement. Even more can be in the R and Quarto files that are part of your project.*

**?@tbl-summarytable** shows a summary of the data.

Note the loading of the data providing a **relative** path using the ../../ notation. (Two dots means a folder up). You never want to specify an **absolute** path like C:\ahandel\myproject\results\ because if you share this with someone, it won’t work for them since they don’t have that path. You can also use the here R package to create paths. See examples of that below.

## 4.2 Basic statistical analysis

*To get some further insight into your data, if reasonable you could compute simple statistics (e.g. simple models with 1 predictor) to look for associations between your outcome(s) and each individual predictor variable. Though note that unless you pre-specified the outcome and main exposure, any “p<0.05 means statistical significance” interpretation is not valid.*

**?@fig-result** shows a scatterplot figure produced by one of the R scripts.

## 4.3 Full analysis

*Use one or several suitable statistical/machine learning methods to analyze your data and to produce meaningful figures, tables, etc. This might again be code that is best placed in one or several separate R scripts that need to be well documented. You want the code to produce figures and data ready for display as tables, and save those. Then you load them here.*

Example **?@tbl-resulttable2** shows a summary of a linear model fit.

# 5. Discussion

## 5.1 Summary and Interpretation

*Summarize what you did, what you found and what it means.*

## 5.2 Strengths and Limitations

*Discuss what you perceive as strengths and limitations of your analysis.*

## 5.3 Conclusions

*What are the main take-home messages?*

*Include citations in your Rmd file using bibtex, the list of references will automatically be placed at the end*

This paper (Leek & Peng, 2015) discusses types of analyses.

These papers (McKay, Ebell, Billings, et al., 2020; McKay, Ebell, Dale, et al., 2020) are good examples of papers published using a fully reproducible setup similar to the one shown in this template.

Note that this cited reference will show up at the end of the document, the reference formatting is determined by the CSL file specified in the YAML header. Many more style files for almost any journal [are available](https://www.zotero.org/styles). You also specify the location of your bibtex reference file in the YAML. You can call your reference file anything you like, I just used the generic word references.bib but giving it a more descriptive name is probably better.

# 6. References

Leek, J. T., & Peng, R. D. (2015). Statistics. What is the question? *Science (New York, N.Y.)*, *347*(6228), 1314–1315. <https://doi.org/10.1126/science.aaa6146>

McKay, B., Ebell, M., Billings, W. Z., Dale, A. P., Shen, Y., & Handel, A. (2020). Associations Between Relative Viral Load at Diagnosis and Influenza A Symptoms and Recovery. *Open Forum Infectious Diseases*, *7*(11), ofaa494. <https://doi.org/10.1093/ofid/ofaa494>

McKay, B., Ebell, M., Dale, A. P., Shen, Y., & Handel, A. (2020). Virulence-mediated infectiousness and activity trade-offs and their impact on transmission potential of influenza patients. *Proceedings. Biological Sciences*, *287*(1927), 20200496. <https://doi.org/10.1098/rspb.2020.0496>