You will be working on a research paper for your final project. This project will include identifying a topic/problem that you want to solve using data science. While the final solution to the problem does not need to be provided via programming – you will be doing some exploratory data analysis, transformations, and summary statistics on the data via R. You are welcome to create a model based on what you have learned in this course to solve the problem, but this is not required. Instead, a recommendation is required for a model or method you would implement to solve the problem. There are 3 steps to this project, with the final deliverable being a formal paper (completed in R Markdown) that outlines the problem, shows the analysis done with the data, and concludes with your recommendation for next steps. Each step provides detailed information that you must include in each phase of the project.

Identify a topic or a problem that you want to research. Provide an introduction that explains the problem statement or topic you are addressing. Why would someone be interested in this? How is it a data science problem?

```
I would like to research heart disease. Heart disease is one of the leading causes of death among humans
```

Draft 5-10 Research questions that focus on the problem statement/topic.

```
1. Are people with diabetes more likely to develop heart disease at an early age?
2. Is there a certain demographic group that is at higher risk of heart disease than others?
3. Does frequent exercise lower the risk of developing heart disease?
4. Are smokers more likely to develop heart disease?
5. How many deaths does heart disease cause?
```

Provide a concise explanation of how you plan to address this problem statement.

```
We can look through datasets that have other medical data so that we can see what the most likely predi
```

Discuss how your proposed approach will address (fully or partially) this problem.

```
If we find out that everyone with heart disease has a certain feature in common, we can start educating
```

Do some digging and find at least 3 datasets that you can use to address the issue. (There is not a required number of fields or rows for these datasets) Original source where the data was obtained is cited and, if possible, hyperlinked.

```
[kaggle study](https://www.kaggle.com/datasets/johnsmith88/heart-disease-dataset)

@Manual{johnsmith,
  title = {knitr: Heart disease dataset},
  author = {David Lapp},
  year = {1988},
  url = {https://www.kaggle.com/datasets/johnsmith88/heart-disease-dataset},
}

[kaggle study2](https://www.kaggle.com/datasets/fedesoriano/heart-failure-prediction)

@Manual{fedesoriano,
```

```
  title = {knitr: Heart Failure Prediction dataset},
  author = {fedesoriano},
  year = {2021},
  url = {https://www.kaggle.com/datasets/fedesoriano/heart-failure-prediction},
}

[kaggle study3](https://www.kaggle.com/datasets/belsonraja/heart-disease-prediction)

@Manual{belsonraja,
  title = {knitr: Heart disease prediction},
  author = {Belsonraja},
  year = {2022},
  url = {https://www.kaggle.com/datasets/belsonraja/heart-disease-prediction},
}
```

Source data is thoroughly explained (i.e. what was the original purpose of the data, when was it collected, how many variables did the original have, explain any peculiarities of the source data such as how missing values are recorded, or how data was imputed, etc.). Identify the packages that are needed for your project.

```
https://www.kaggle.com/datasets/johnsmith88/heart-disease-dataset
This data set dates from 1988 and consists of four databases: Cleveland, Hungary, Switzerland, and Long

https://www.kaggle.com/datasets/fedesoriano/heart-failure-prediction
This dataset was created by combining different datasets already available independently but not combin

https://www.kaggle.com/datasets/belsonraja/heart-disease-prediction
Its the heart disease dataset from this dataset we can derive various insights that help us know the we

#all data is numerical and none appears to be missing.
packages needed: knitr, ggplot, purr, tidyr, tidyverse, stats, car, dplyr
```

What types of plots and tables will help you to illustrate the findings to your research questions?

```
table()
scatterplot()
density plot
geom_boxplot
geom_bar
pandoc table
```

What do you not know how to do right now that you need to learn to answer your research questions?

```
I need to learn how to better read the summary function to draw conclusions on datasets
```

You can use the following template for Step 1: Introduction Research questions Approach How your approach addresses (fully or partially) the problem. Data (Minimum of 3 Datasets - but no requirement on number of fields or rows) Required Packages Plots and Table Needs Questions for future steps

```
knitr::write_bib(c("knitr", "stringr"), "", width = 60)
```

```
## @Manual{R-knitr,
##   title = {knitr: A General-Purpose Package for Dynamic
##     Report Generation in R},
##   author = {Yihui Xie},
##   year = {2022},
##   note = {R package version 1.41},
##   url = {https://yihui.org/knitr/},
## }
##
## @Manual{R-stringr,
##   title = {stringr: Simple, Consistent Wrappers for Common
##     String Operations},
##   author = {Hadley Wickham},
##   year = {2022},
##   note = {R package version 1.5.0},
##   url = {https://CRAN.R-project.org/package=stringr},
## }
##
## @Book{knitr2015,
##   title = {Dynamic Documents with {R} and knitr},
##   author = {Yihui Xie},
##   publisher = {Chapman and Hall/CRC},
##   address = {Boca Raton, Florida},
##   year = {2015},
##   edition = {2nd},
##   note = {ISBN 978-1498716963},
##   url = {https://yihui.org/knitr/},
## }
##
## @InCollection{knitr2014,
##   booktitle = {Implementing Reproducible Computational
##     Research},
##   editor = {Victoria Stodden and Friedrich Leisch and Roger
##     D. Peng},
##   title = {knitr: A Comprehensive Tool for Reproducible
##     Research in {R}},
##   author = {Yihui Xie},
##   publisher = {Chapman and Hall/CRC},
##   year = {2014},
##   note = {ISBN 978-1466561595},
##   url =
##     {http://www.crcpress.com/product/isbn/9781466561595},
## }
```