MADA Data Analysis Project

Monica Chan

2021-09-17

The structure below is a possible setup for a data analysis project (including the course project). For a manuscript, adjust as needed.

# 1 Summary/Abstract

*Write a summary of your project.*  
Applying MADA learned techniques to a RT-PCR dataset to determine the effectiveness of sample enhancement strategies for environmental surface sampling of low bio-burden viruses.

# 2 Introduction

## 2.1 General Background Information

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

### 2.1.1 Surfaces Preparation

Stainless-steel coupons (10x20in) were triple washed with detergent, rinsed with RO water, sprayed with 70% Ethanol and left to air dry. The dry coupons were wrapped in craft paper and autoclaved on a solid cycle. Before use, the coupons are unwrapped while in a BSC

### 2.1.2 Inoculum Preparation

A serial dilution of heat-inactivated SARS-CoV-2 (~10^6 copies/uL, ATCC Strain) and Viral Transport Media (CDC protocol) was created

### 2.1.3 Inoculation of Surfaces

## 2.2 Description of data and data source

*Describe what the data is, what it contains, where it is from, etc. Eventually this might be part of a methods section.*

Data was self recorded over several experimental runs of an on going series. The current set has 379 observations and 23 variables. Additional data will be added into this set as experiments are completed throughout the course. The data itself is a combination of observed processing variables and the resulting RT-PCR data. Copy data may eventually be included. It is currently calculated in an excel document, but may be reproduced in this project for best practice). A copy of the data has been placed in the data folder of this repository.

## 2.3 Questions/Hypotheses to be addressed

There are 3 major aims of this:

1. Do the sample enhancement strategies work?
2. How well do they work compared to each other?
3. Are there any environmental variables recorded that have a positive impact on detection?

Testing several sample enhancement strategies (Nanoparticles and Bait & Capture) for molecular detection of heat-inactivated SARS-CoV-2 from stainless-steel surfaces. The primary aim of this work is to determine if any of the enhancement strategies increase the concentration of RNA copies in a sample that will express itself as a lower Ct value from the Rt-PCR output. Lower Ct values equate to more copies being present in the sample meaning less time needed to confirm presence.

In the case the sample enhancements work, we would then compare how well they worked against an automated extraction method. The degree of “wellness” will be determined by how low the Ct value results are produced after a sample has going through a specific method.

In addition to the comparison of enhancement strategies the effects of the tool used for sampling (macrofoam or knit-fiber), time held, and other environmental variables will be explored any significant effects, with interest in positive impacts to detection.

# 3 Methods and Results

*In most research papers, results and methods are separate. You can combine them here if you find it easier. You are also welcome to structure things such that those are separate sections.*

## 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.*

## 3.3 Exploratory analysis

*Use a combination of text/tables/figures to explore and describe your data. You should produce plots or tables or other summary quantities for the most interesting/important quantities in your data. Depending on the total number of variables in your dataset, explore all or some of the others. FIgures produced here might be histograms or density plots, correlation plots, etc. Tables might summarize your data.*

*Continue by creating plots or tables of the outcome(s) of interest and the predictor/exposure/input variables you are most interested in. If your dataset is small, you can do that for all variables. Plots produced here can be scatterplots, boxplots, violinplots, etc. Tables can be simple 2x2 tables or larger ones.*

*To get some further insight into your data, if reasonable you could compute simple statistics (e.g. t-tests, simple regression model with 1 predictor, etc.) 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.*

Table 3.1 shows a table summarizing the data.

Table 3.1: Data summary table.

|  |  |  |
| --- | --- | --- |
|  | Height | Weight |
| Min. | 133.00 | 45.00 |
| 1st Qu. | 155.25 | 54.25 |
| Median | 166.00 | 73.00 |
| Mean | 165.50 | 72.00 |
| 3rd Qu. | 177.25 | 87.50 |
| Max. | 192.00 | 110.00 |

Figure 3.1 shows a scatterplot figure produced by one of the R scripts.

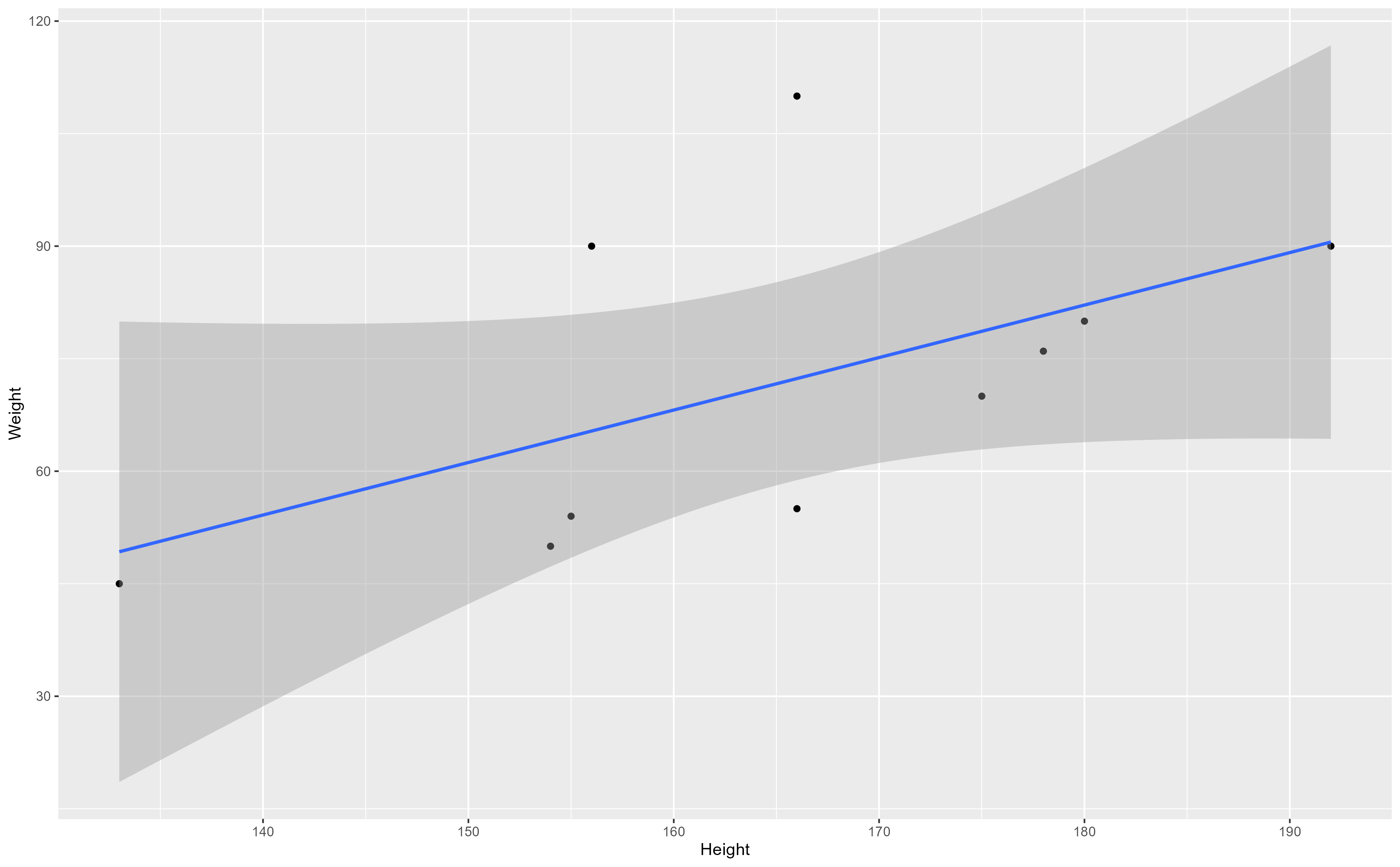


Figure 3.1: Analysis figure.

## 3.4 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 table 3.2 shows a table summarizing a linear model fit.

Table 3.2: Linear model fit table.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| term | estimate | std.error | statistic | p.value |
| (Intercept) | -43.7883068 | 61.1150617 | -0.7164896 | 0.4940713 |
| Height | 0.6996272 | 0.3675692 | 1.9033889 | 0.0934786 |

# 4 Discussion

## 4.1 Summary and Interpretation

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

## 4.2 Strengths and Limitations

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

## 4.3 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.

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.

# References

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