E. coli resistance patterns from Community Resistance in Athens Project

Andreas Handel

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.*

# 2 Introduction

## 2.1 General Background Information

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

As microbes are evolving to resist the drugs we commonly use to treat infections, antibiotic resistance is becoming a rapidly growing health threat worldwide.Many key AR threats, can colonize the human intestinal tract, and intestinal colonization is a well-established risk factor for AR infection. The Community Resistance in Athens Project seeks to better understand carriage rates of multi-drug resistant *E. coli* such as extended spectrum beta lactamase-producing Enterobacteriaceae and carbapenem resistant Enterobacteriaceae. In addition, the CRAP project is gathering data on the resistance patterns of *E. coli* from individuals in Athens. The data I will be using is a combination of microbiological data that I have collected and the epidemiological data from the survey. The microbiological data quantifies the resistance patterns of *E. coli* from individuals to different antibiotics. The survey queries the study participants on a variety of personal and lifestyle questions, including drinking water sources, diet, economic information, etc.

## 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.*

As lab technician for the CRAP project, I have collected the data on the E. coli resistance patterns myself. From participants, we collect a stool sample, take a small diluted aliquot, and grow its E. coli on agar that selects for E. coli. I further purify these isolates, then test 48 or less from each sample, provided E. coli grows, onto different agar plates containing either ampicillin, tetracycline, ceftriaxone, ciprofloxacin, or trimethoprim. I record the patterns for each participant and antibiotic. The data on food consumption, animal contact, and the other epidemiological data comes from a survey each of the participants must take as a part of the study. This survey is IRB approved and survey and sample collection are organized by the Clinical and Translational Research Unit at the University of Georgia.

We have data from more than 350 individuals at this points, and the resistance patterns of approximately 13,000 E. coli isolates. We have already made figures for basic questions concerning this data, like what proportions of the study population that show resistance to each antibiotic type and demographics, so I would like to focus on environmental factors.

## 2.3 Questions/Hypotheses to be addressed

*State the research questions you plan to answer with this analysis.*  
For this project, we have already observed the general resistance patterns and carriage rates of *E. coli*, so I would like to see if there is a relationship between the resistance patterns and the diet of the individuals. Some researchers presume that commensal gut bacteria act as reservoirs for genes with antibiotic resistance, which could lead to horizontal gene transfer and possibly the development of dangerous multi-drug resistant bacteria (Mathur, et. al). There is abundant research on antibiotic resistance exposure in healthcare settings, and there is a gap in knowledge on everyday exposures, like food consumption types. There is evidence that bacteria from fermented dairy and meat commonly carry AR gene (Mathur, et.al). In addition, according to the CDC, raw fruits and vegetables may be contaminated by soil and water (CDC 2020). With respect to this information, I have a few questions.

* Do individuals who eat meat more than half of the days in the last week have higher resistance than do those who do not?
* Do individuals who eat dairy more than half of the days in the last week have higher resistance than do those who do not?
* Do individuals who eat raw fruits or vegetables more than half of the days in the last week have higher resistance than do those who do not?

In addition to these questions, I am interested in the following, if I have time:

* Are resistance patterns influenced temporally?
* Are individuals with exposure to animals (pets, cattle, etc.) have higher resistance patterns than those who do not?
* In there any relationship between resistance of *E. coli* in an individual and the age of the participant?

# 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.* This is partly explained in the “Description of data and data source” section. In addition, the resistance patterns will be in a single csv file that I create from an excel sheet. The information on food consumption has been cleaned by another member of the CRAP team and he will share the information with my promptly. It will also be in a csv file.

## 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.

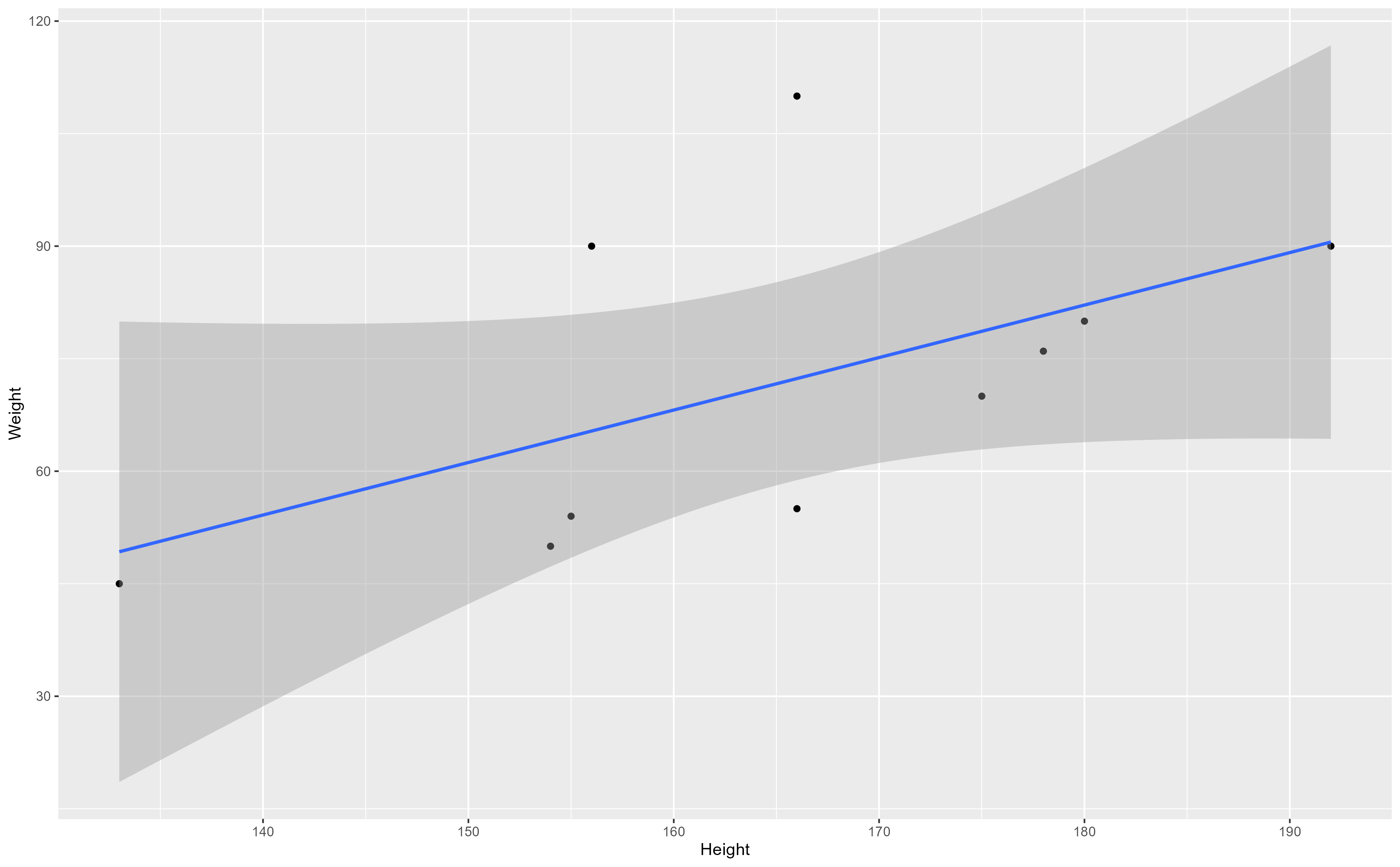


Figure 3.1: Analysis figure.

## 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.

# 5 References

(I will put these links in proper format soon.)

* <https://www.cambridge.org/core/journals/epidemiology-and-infection/article/risk-factors-for-antibioticresistant-e-coli-in-children-in-a-rural-area/D3A8BA423024A18D592BAA70719608A0>
* <https://www.sciencedirect.com/science/article/pii/S1473309918302962?casa_token=j_ZSlMulgkcAAAAA:dUmvDkbjCNWA2ELFH1tBujfCqNQVtCEWm9aGJOaYCzZ8aQDgWEOie1US-SLNsw3AqNA9xbW2nQ>
* <https://www.sciencedirect.com/science/article/pii/S1473309918302962?casa_token=j_ZSlMulgkcAAAAA:dUmvDkbjCNWA2ELFH1tBujfCqNQVtCEWm9aGJOaYCzZ8aQDgWEOie1US-SLNsw3AqNA9xbW2nQ>
* <https://www.nature.com/articles/39767>
* <https://www.sciencedirect.com/science/article/pii/S0168160505002618>
* <https://www.cdc.gov/foodsafety/challenges/antibiotic-resistance.html>

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>