Group 3: Data Analysis

Nick Cummings, Liza Nicoll, Emily Ramos, and Yiding Zhang

University of Massachusetts, Amherst

April 26, 2014

Overview

1. Group Analysis
   - Introduction
   - Summary of the Dataset
   - Initial Model
2. Individual Analysis
   - Regression Trees
   - Resampling Inference or something
   - Not sure what Yiding decided on
   - Model Selection

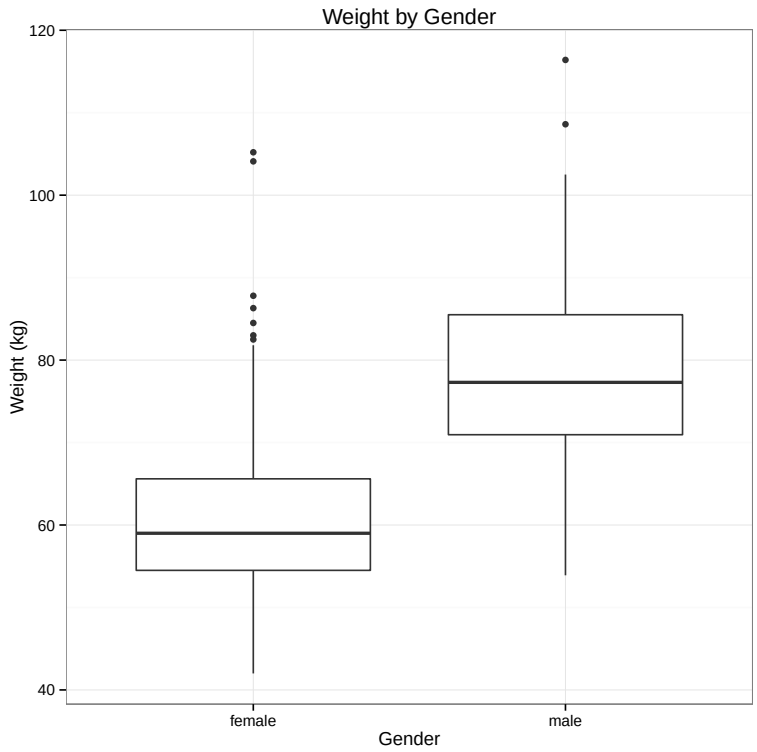
Introduction

Describe dataset

- introduce the dataset, give the reason data was originally collected
- describe sample used
- basically introduce the project and discuss the usefulness of and applications for this data

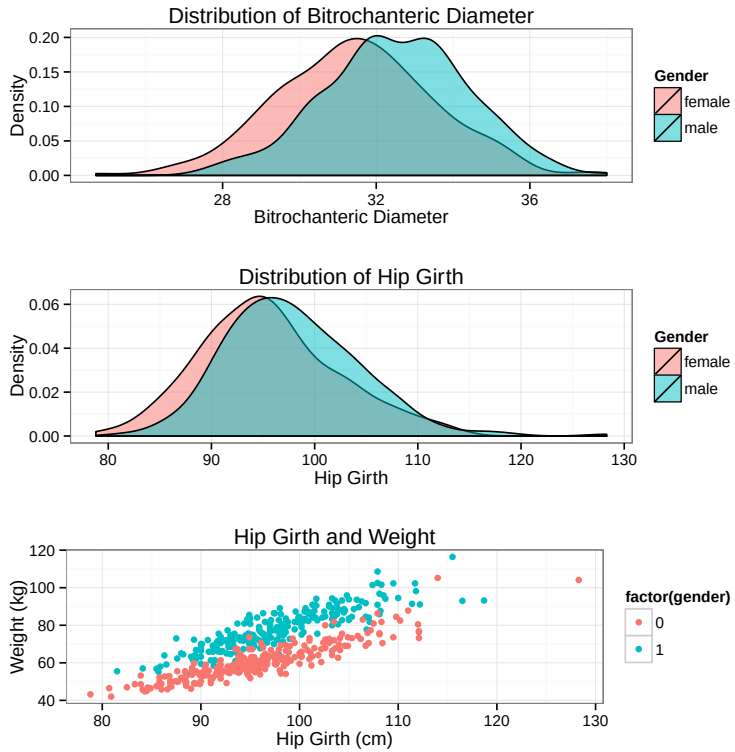
Weight

Weight was measured for 507 physically active individuals - 247 men and 260 women. The distribution ranged from 42 kilograms to 116.4 kilograms. The mean weight and quartiles was signifigantly higher for men than women. We observe several outliers in the upper end of the range for both men and women. This may be due to the fact that the population sampled included a number of highly physicaly fit individuals with higher than average muscle mass.



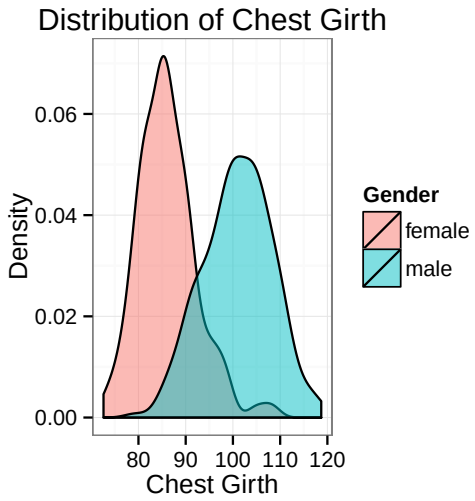
Bitrochanteric Diameter and Hip Girth

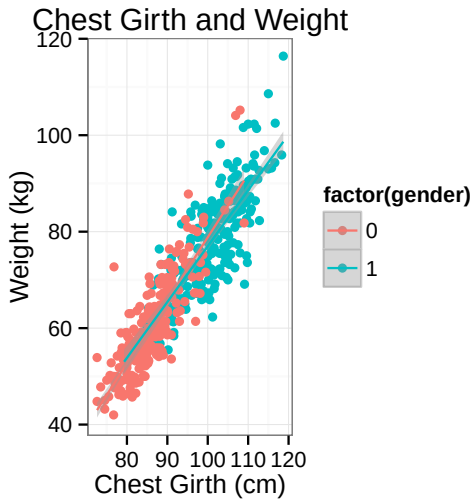
Bitrochanteric diameter is the distance between the outer points of the hips and hip girth is the circumference of the hip area measured at the level of the bitrochanteric diameter. The density distributions for both measures are normally distributed (though hip girth is skewed slightly right) and very similar in distribution for both men and women. A scatter plot of hip girth vs. weight suggests that weight increases linerally with increase in hip girth.

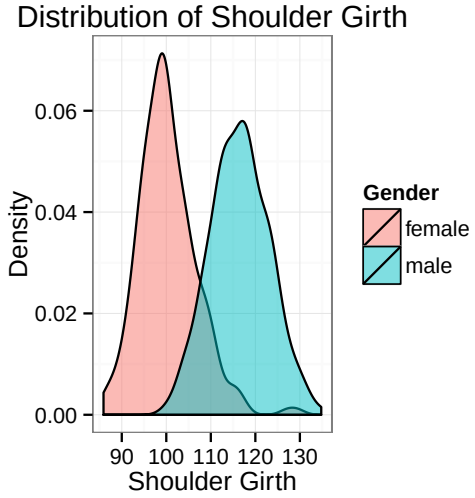


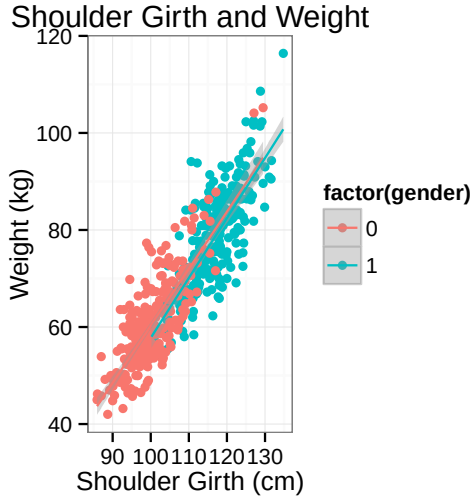
Chest and Shoulder

Chest girth was measured at the nipple line in males and just above breast tissue in females at mid-expiration and shoulder girth was measured over deltoid muscles in both males and females. The density distributions for the two variables are quite similar. Women have narrower, though slightly skeewed, distribution with a much lower mean than that of the men. The scatterplots are also similar in that the regression lines for men and women are nearly identical, indicating that weight increases linearly with increase in shoulder girth, independant of gender.



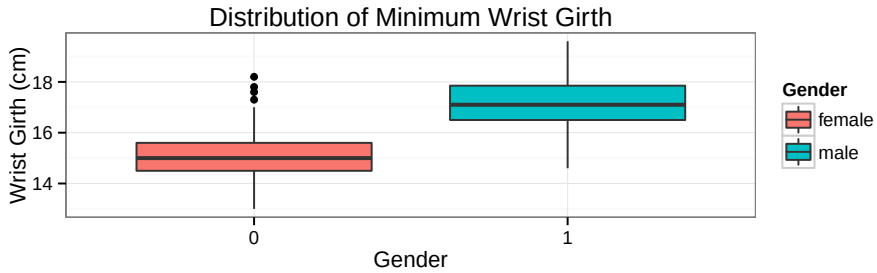


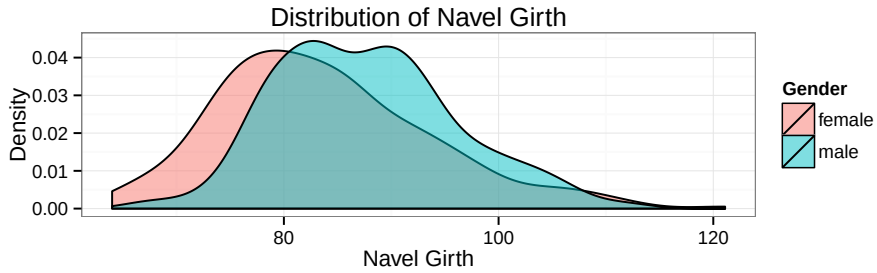


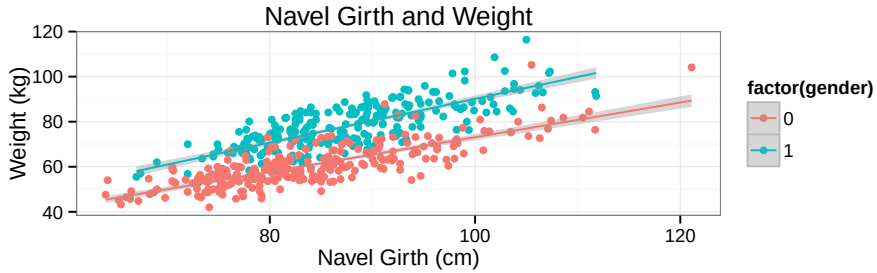


Wrist and Navel

Wrist minimum girth is an average of right and left girths and navel (or abdominal) girth was measured at umbilicus and the iliac crest, using the iliac crest as a landmark. Wrist girth is bimodally distributed, but when divided into male and female, the distributions are normal with some outliers at the high end of the range for females. The distributions for navel girth is normal and remarkably similar for males and females. The scatterplot of navel girth against weight shows a linear relationship with weight increasing with increased naval girth.







Multiple Linear Regression Model

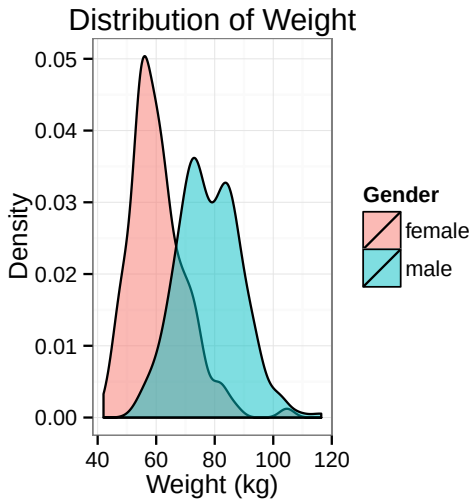
Text describing why this model

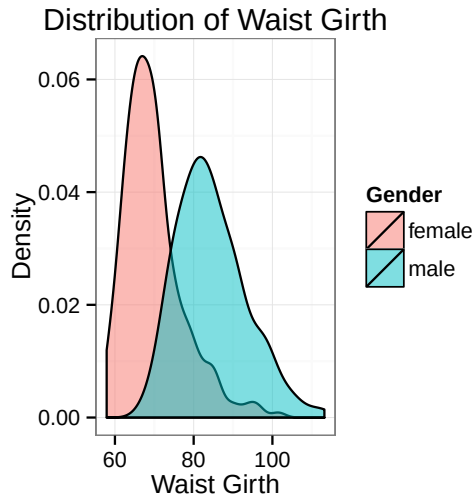
Model

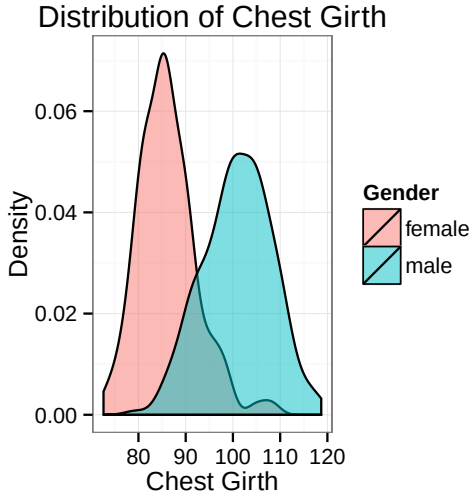
Present predictive model for weight (2) in paper
-fit this model and discuss results
-discuss potential uses for models of this data (finding ideal weight based on skeletal measurements?) and potential problems for applying to whole population (all participants were physically fit)

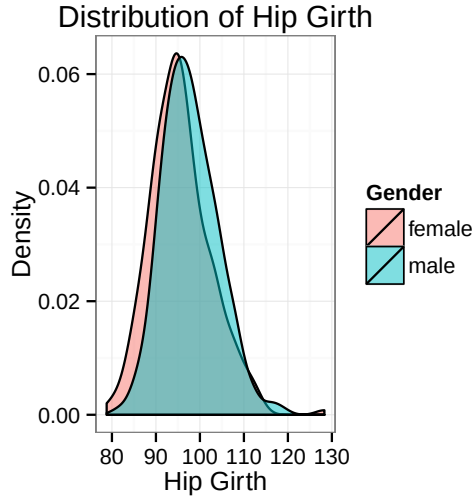
Differences between Males and Females (Liza)

- Are there significant differences in the body measurements most useful for predicting weight in males and females?
- Is one regression formula appropriate for predicting weight for both genders?
- Can we use regression trees to help explore these questions?



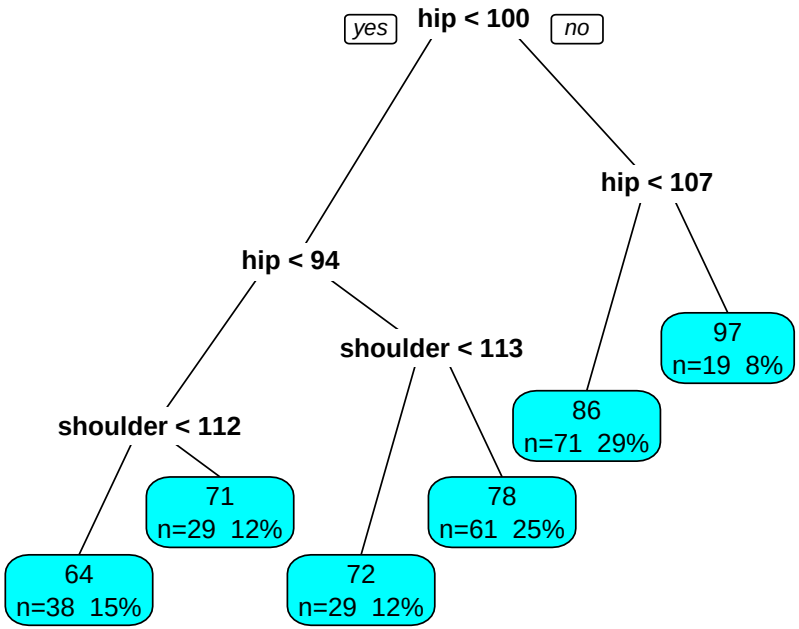




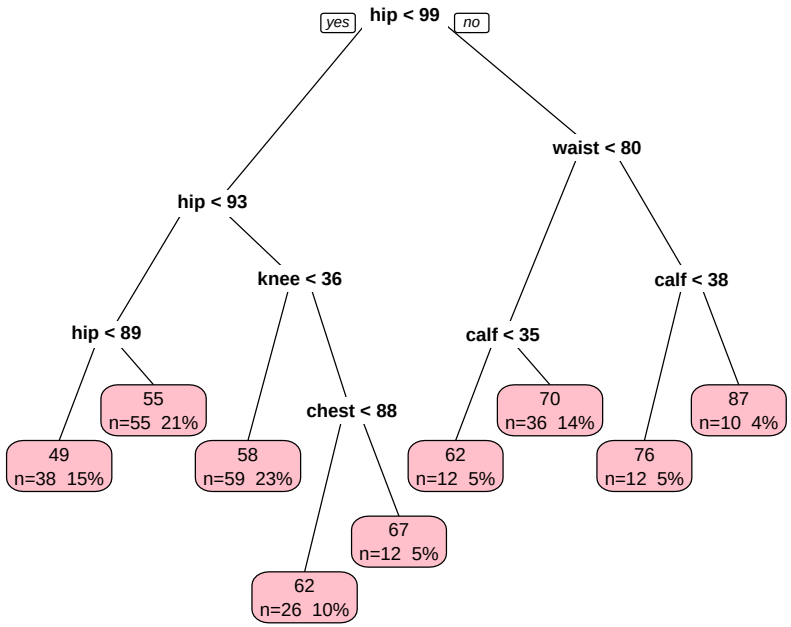


Regression Trees

Pruned Tree, Weight (Male)



Regression Tree, Weight (Female)



- Variables used in male tree: hip shoulder girths
- Variable used in female tree: hip, knee, chest, waist and calf girths

Conclusions (Liza)

- Regression trees are useful for exploring data and provide a useful alternative to parametric regression methods, though are not intended for making predictions.
- Results here suggest that separate models for males and females might be appropriate.
- Model fitting and selection exercises could test this hypothesis.

Nick

Nick

Yiding

Yiding

Emily

Emily