What can body composition data tell us?
How are diffeerent measurements related?

Michaela Bayerlova *Washington State University*

In this article I compare some self collected *body composition measurement data* with one another and try to gather interesting findings, so that I can drive some conclusions and possibly generalizations. The data gets cleaned up first to eliminate the noisy data and make the results more significant. Then the analytics comes in. For all statistical and data analytical reference, see the Appendix section 6

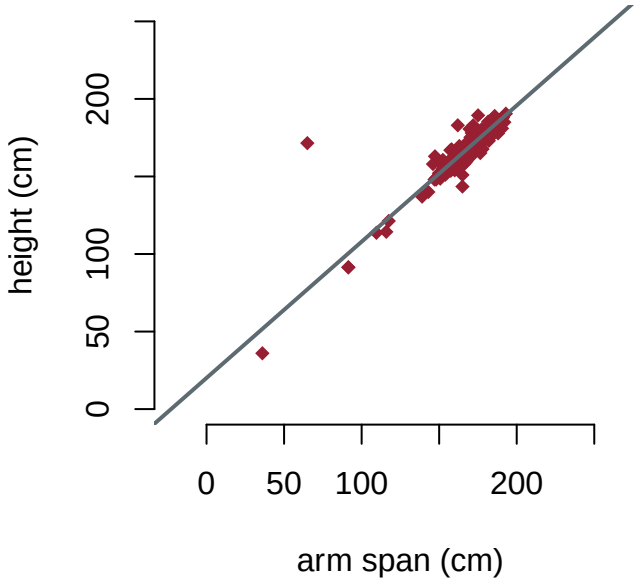
Keywords: Data Cleanup; Data Analysis; Data Visualization

November 22, 2020

1 Introduction

The early beginning of statistics and data analytics started with taking data and observing it. This project involves all these steps as well. Students collected human body measurement data themselves and then analyzed it based on their personal interest.

Figure 1: The relationship between human height and arm span!



The one graphic is a nice representation of two data variable's relationship. The variables are height and arm span. The observation is an almost perfect linear relationship with same proportions or measurements.

This graphic is also included in the main analysis of the data and one point of interest that has been analyzed.

(a) Height versus Head Height plot on cleaned measure data.

(b) Height versus Head Height plot on the Instructor's cleaned measure data.

Figure 2: Height versus Head Height on different cleaned data sets

The Two In One Graphic has 2 graphs of height versus arm span next to each other. The left hand side graphic is the one you see as the OneGraphic above as well. It is the graph from my self-cleaned data, whereas the data of second one is the Instructor cleaned data. I put them next to each other so that it is able to compare them. It is possible to see that they follow the same shape and layout.

2 Research Question: How do body composition ratios relate to gender and age?

2.1 How many average 'head height' lengths are males and females relative to their respective average 'height'?

2.2 Does grouping with ratios of 'head height' and 'height' for males and females match with groupings based on age?

2.3 How does height correspond to arm span?

3 Data Description

The data has been collected for a university class project, STATS 419 at Washington State University. Each student collected the data on oneself and other 9 people in the best case. The measurements were data_collector, person_id, side, height, head.height, head.circumference, hand.length (left, right), hand.width (left, right), hand.elbow (left, right), elbow.armpit (left, right), arm.reach (left, right), arm.span, foot.length (left, right), floor.kneepit (left, right), floor.hip (left, right), floor.navel and floor.armpit (left, right). Additional information like units, writing, eye, eye_color, swinging, age, gender, quality, minutes, ethnicity and notes were also collected. More details about the data collection process were also identified and clarified in a handout, that each student created to give their participants. The students had to do this at their own time. Due to a global pandemic of Covid-19 social distancing was needed to be followed and the data was collected by the individuals themselves with measuring tape.

Then the professor created a large file of all the measurements together and gave it back to us students. Now it is time to use this primary data for data analysis. Each student had to come with research questions that they want to study and observe. My personal research questions are: How do body composition ratios relate to gender and age? -> How many average 'head height' lengths are males and females relative to their respective average 'height'? -> Does grouping with ratios of 'head height' and 'height' for males and females match with groupings based on age? -> How does height relate to arm span?

For more details, reference the Appendix section.

3.1 Summary of Sample

The sample we found initially had a lot of noise, which required a good clean up. The findings were some duplicate data, wrong measurements, different notations and units and some outliers. All of these errors have been taken into account in the clean up section. The data left is much more less noisy and accurate to do proper Data Analysis.

3.2 Summary Statistics of Data

The summary statistics of the collected data includes means and standard deviations of all variables. It also shows the quartiles and therefore how the data is distributed. Another summary statistic can be the dimensions of the data. The initial data set had 428 observations and 37 variables. After the instructors clean up, there were only 251 left.

4 Key Findings

The key findings from the analysis are giving the answers to my research questions. The main research question was asking: How do body composition ratios relate to gender and age. I only started looking at the head height vs height proportions of every person. Then the male and female data was separated so that I can find the average of each. The next key finding was, that the data grouping by head height and height proportions or height and age do give the same result for ~90 % of the time. This means, that the grouping comparisons are valid and work well for this data. An addition to this is the interest of what percentage is considered heroic. There I was surprised that it was ~15 % of the whole data when grouping based on proportions. It seemed quite a lot to me. And out of those 15 %, 61.54 % have been male. Another grouping by internet vales of average height showed, that ~53 % of categorized 'heroic' people this time were male.

A very great finding was also, that height and arm span have a strong linear relationship. The taller one person is, the wider is the corresponding arm span. Another interesting fact about it is, that the measurements of height and arm span of a person are almost identical. When doing a proportion again, and then only letting values in between 0.9 and 1.1 (a 90% match) being true, then ~89 % of the data is applicable.

When coming back to my primary research question, I can say, that the body composition ratio of head height vs height do relate to gender and more to age. It relates more to age, because a child's ratio differs from the one of an adult. And there a clear line can be seen. Whereas, with gender there is not much of a possible distinguish, because the ratio of an average of 8 is the same.

5 Conclusion

As a conclusion I can say, that I have been able to find very interesting results, that let me explain my research questions. I can for sure say, that there is a lot more to learn from body composition data.

Table 1: Descriptive Statistics and Correlation Analysis

	M	SD	1	2
1 Height (cm)	166.9	29.75	1	
2 Arm Span (cm)	210.7	35.58	.90***	1
3 Head Height (cm)	22.2	3.56	.66***	.66***

Notes: Pearson pairwise correlations are reported;
a two-side test was performed to report correlation significance.

$^{\dagger}p < .10$ $^{*}p < .05$ $^{**}p < .01$ $^{***}p < .001$

The table shows, the mean and standard deviation for all of the variables and then the correlation significance factor between those variables. We see from there, that the correlation between arm span and height is very strong and positive (0.9). So when someone is taller than someone else, that person's arm span should also be bigger. The relation between head height and height is 0.66, the same is true for head height and arm span. The correlation is only average and cannot be taken as very significant.

6 APPENDICES

6.1 Data Provenance

Data Provenance is very crucial and important in this project and for every data collection process and data analytics evaluation. The data collection is one of the most important steps in data analytics, it requires authenticity, enabling trust into the data and allowing reproducibility. The origin of the data are students in a university statistics class, that took the body measurements themselves. All of the data then was combined into one text file by the instructor. The validation of the initial data turns out to be not as trust worthy as expected. There have been evaluated some duplicate and wrong data which have been cleaned up in a self clean up process and also by the instructor, who later on provided a cleaned data set in order to be sure of data provenance and having data to trust.

The data is reproducible. To reproduce the data a handout has been created. It is attached below for reference.

6.1.1 Data Collection Handout

Figure 3: Handout Page 1

MEASURE PROJECT

Instructions:

Please, fill out the tables below to the best of your abilities. They contain some information (covariates) and then it will require to take a measuring tape and maybe grab another person for assistance. The data that will be collected will be safe and in later steps anonymously used for analysis.

I provide definitions of the measurements, pictures for visualizing and examples of my data.

Covariates:

Name	Units	Writing	Eye	Eye color	Swinging	Age	Gender	Quality	Minutes	Ethnicity	Notes
MichaelaB	cm	right	right	brown	right	21	female	10	23	White	White European

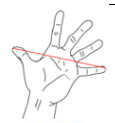
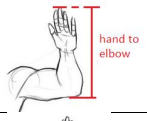
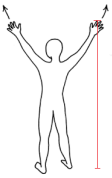
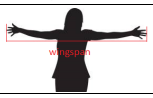
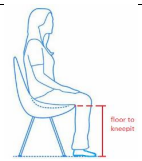
Definitions:
Units: Units of the measurements
Writing: With what hand do you write?
Eye: What is your dominant eye?
Swinging: What hand do you throw a ball or swing a racket?
Quality: What is the quality of your measurements?
Minutes: How long did it take you to complete?
Notes: Any additional information or clarification?
Eye color, Age, Gender, Ethnicity: self-explanatory

Measurements:

What to measure	Measurements		Example picture for clarification
Height	175.6		
Head.height	21.0		
Head.circumference	53.8		
Hand.length (left) Hand.length (right)	18.5 18.3		

Michaela Bayerlova
michaela.bayerlova@wsu.edu
011594781

Figure 4: Handout Page 2

Hand.width (left) Hand.width (right)	21.2 20.5		
Hand.elbow (left) Hand.elbow (right)	47.3 47.0		
Elbow.armpit (left) Elbow.armpit (right)	31.5 31.5		
Arm.reach (left) Arm.reach (right)	221.0 223.0		
Arm.span	176.2		
Foot.length (left) Foot.length (right)	26.4 26.3		
Floor.kneepit (left) Floor.kneepit (right)	45.0 45.2		
Floor.hip (left) Floor.hip (right)	94.5 95.0		
Floor.navel	109.0		
Floor.armpit (left) Floor.armpit (right)	141.0 141.0		

Michaela Bayerlova
michaela.bayerlova@wsu.edu
011594781

(a) Thomas et al. (2020) discuss this.

(b) Schnitt realer Sensor (Thomas et al. 2020)

Figure 5: Der Sensor in Theorie und Verwirklichung.

6.2 Preparing the Report Workspace

Below is the necessary functions and libraries required to run the code referenced in this document.

```
library(devtools);       # required for source_url

path.humanVerseWSU = "https://raw.githubusercontent.com/MonteShaffer/humanVerseWSU/"
source_url( paste0(path.humanVerseWSU,"master/misc/functions-project-measure.R") );

path.functions = ""
```

Below is the code to load the data and prepare it for analysis.

```
path.project = "C:/Users/michaela.bayerlova/Documents/STATS 419/measurement_project/michaela/project-mea

path.to.secret = "C:/_git_/SECRETS/";

file.measure = paste0(path.to.secret,"measure-students.txt");
file.measure.clean = paste0(path.to.secret,"cm.final.measure.txt");

path.github = "https://raw.githubusercontent.com/MichaelaB1/WSU_STATS419_FALL2020/";
source_url( paste0(path.github,"master/Project1/michaela_project_measure/functions/functions_project_mea

# this is your function
# put in the same "units"
# merge left/right
# build proportion data
# and so on ...
# measure.df = prepareMeasureData(measure);
```

6.3 Self Data Cleanup

Create data frame measure At first the data is read into the program and then I save the original data in a data frame called measure.raw. Another data frame is created as well, measure.clean, which is going to be modified and cleaned. To get an initial basic idea and overview about the data, I view the dimensions and summary statistics of the data set.

Remove Duplicates As the next step, the duplicates function is given to use by the instructor to use.

Average values The initial collected data has some entries that have one measurement like height, but then there are also measurements of left and right side, for example the hand length. For data points that have a right and left measurements, the average will be taken in this function. When there is only one measurement, that one will be taken. This enables to get a uniform set of data without missing values for those entries. These new average values are added in new columns to the end of the data frame.

Leave only average values This section is cleaning up the data frame from above. The single entries of average values will be placed instead of the left and right data, which are removed. For further analysis the 'notes' column is not needed and is therefore removed as well.

Unity among gender notation In the collected data, there have been used different notations to indicate 'female' and 'male'. This will cause problems in further analyses and it therefore needs to be adjusted to an overall standard. The standard will be the one provided above. First of all, the whole string entries in the column 'gender' will be converted to lower case characters. Then the wrong assigned data will be newly assigned to 'female' or 'male'.

Converting units to cm Since not all data has been collected in the same units, this has to be fixed in order to compare the data properly. Units used are inches and cm. I have decided to go with cm, so I convert inches to cm, where 1 inch equals to 2.54 cm. Before applying the math there, I have to also change the notation of units of inches. There have been used 'inches', 'Inch', 'in' and 'inch'. The last notation is the preferred one to use. After this change has been made, the unit conversion takes place.

Remove wrong data This step is removing data that cannot be true. For example observations of the data showed, that some arm.reach data point are smaller than the normal height. This cannot be true, it is a wrong measurement and needs to be removed. So the code compares these 2 columns, when a comparison is false the entry will be removed from the data frame.

Remove Outliers This step enables to clean up the data even more such that it has only accurate and reasonable data. The method used uses quantiles and IQR. The data has to be free of NA values first.

Findings: Clean up The clean up removed the main noisy and wrong data, so that the further analysis is taken upon good data to trust and also to trust the analysis results.

Below is the self data clean up function that has all the steps inside of them.

```
measure.clean = prepareMeasureData(file.measure);
# measure.clean;
```

6.4 Instructor Data Cleanup

The instructor has provided another cleaned data set to the students after some issues have arised.

Therefore, only a new text file needs to be loaded and read in as a data frame. Changing gender labels and dropping unnecessary columns as well to obtain the same data frame as I do with my clean up. (no notes column also)

```
measure.clean2 = readDataIntoDataFrame(file.measure.clean);

measure.clean2$gender[measure.clean2$gender == 'F'] <- 'female';
measure.clean2$gender[measure.clean2$gender == 'M'] <- 'male';
# measure.clean2;

# determine if this separation above will give the same as separating by head proportions
measure.clean2 = measure.clean2[,-(18), drop=FALSE];
measure.clean2 = measure.clean2[,-(27:34), drop=FALSE];

# measure.clean2;
```

6.5 Data Analysis

6.5.1 Proportions of head.height vs whole body height

Create a subset with only important variables for the analysis This step creates a subset of the data, so that only the variables are taken which are needed and used for analysis later. This step can be done even without cleaning up the data before from NA values, this way the rows with empty values in other columns do not affect the data subset here, because it still can be used. And the missing values only in these columns can be removed later. The columns left are 'height.NA', 'head.height.NA', 'arm.span.NA', 'age' and 'gender'. The others are dropped.

Create new data frame 'measure.proportions' to work on analysis Right now the NA values from the subset data frame are removed with the na.omit() command.

Assign proportions group This analysis is based on the calculation of how many times the head height fits into the whole body height. Then the proportion values will be grouped into different groups, these are: Group 'F': There the person is most likely a female adult because the proportions are bigger than what kids usually have, which is above 6.5. Group 'M': There the person is most likely a male adult because the proportions are bigger than what kids usually have, which is above 6.5. Group 'K': There the person is most likely a child (male or female) because the proportions are smaller or equal to 6.5.

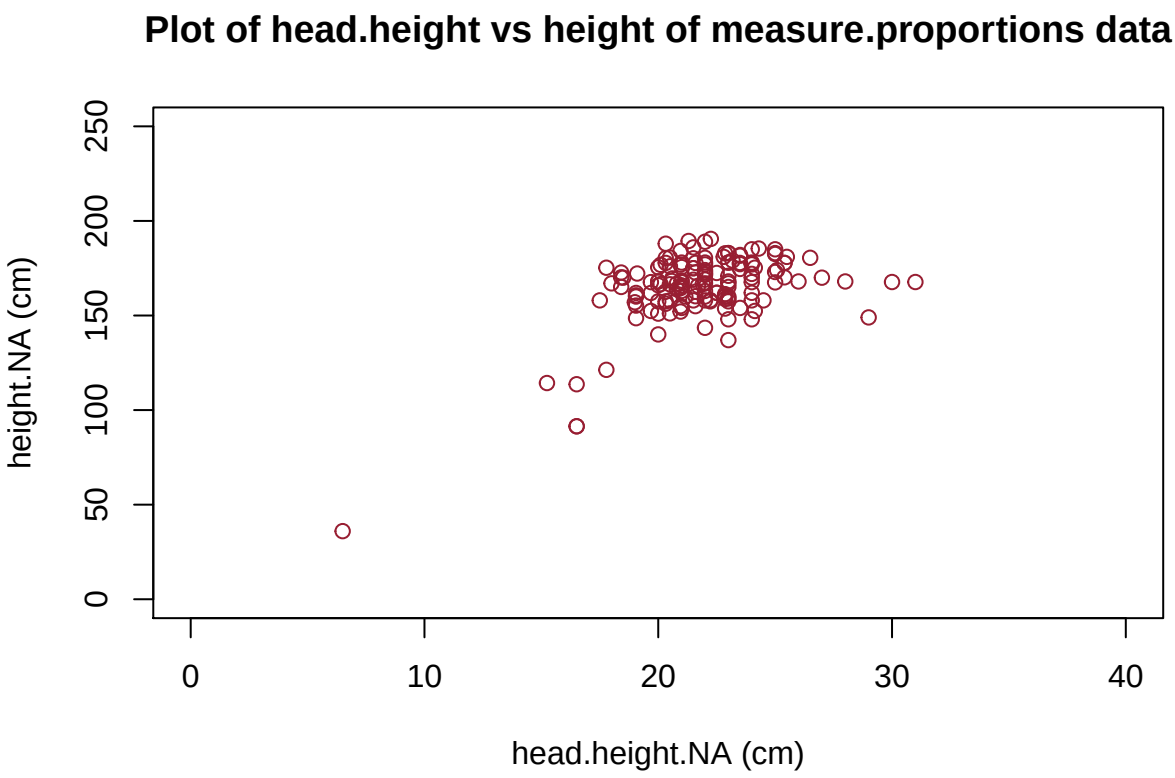
Assign age group This analysis is based on the age and gender of a person. Then these groups are created: Group 'F': There the person is female and 11 years or older. Group 'M': There the person is male and 11 years or older. Group 'K': There the person is a child (male or female) because the age is 10 or younger.

Findings: Assigning proportions and age group The results give three new columns in the data frame. One where the proportions of height/head.height are given, another where these proportions are used to assign each to a label ('F', 'M' or 'K'). These same labels are also assigned in the third column, where the criterion is age. This gained information will then especially be used to compare the 2 assignment methods.

The code for this step is showed below.

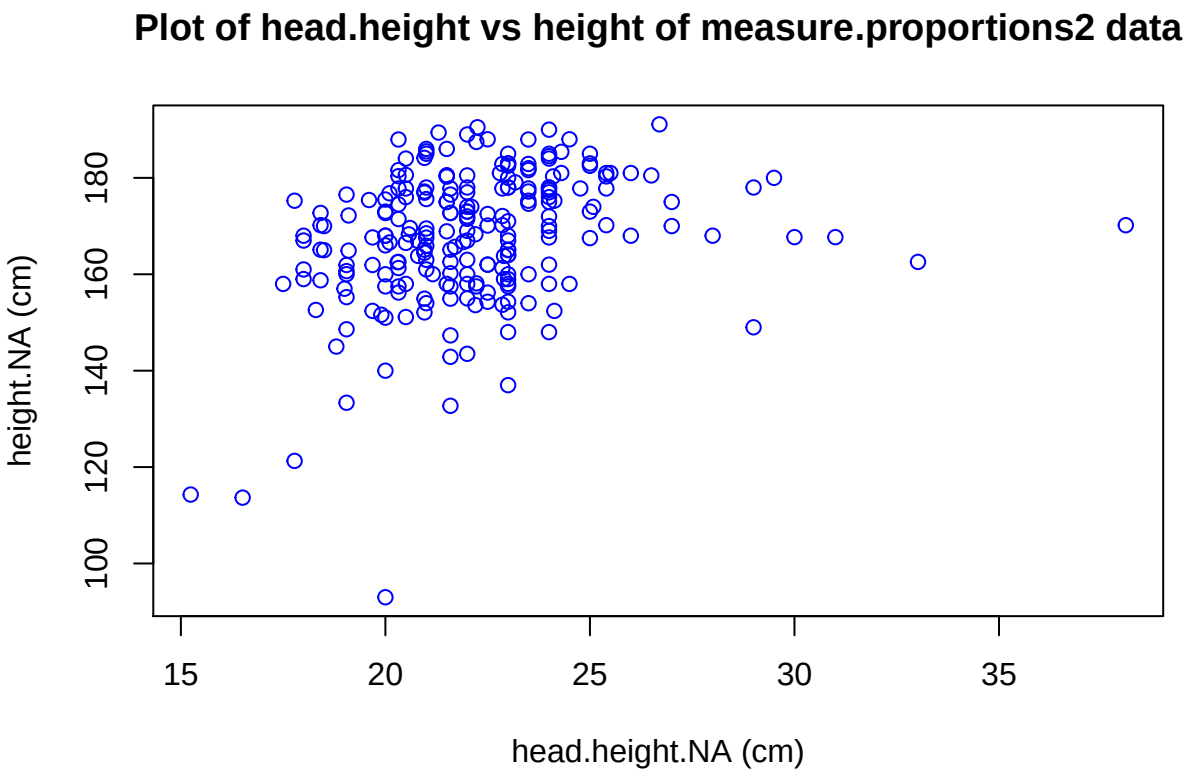
```
# Prepare for analysis of proportions
measure.proportions = prepareProportionsAnalysis(measure.clean);

plot(measure.proportions$head.height.NA, measure.proportions$height.NA, col = "#981e32",main = paste("P
```




```
# same step for instructors data
measure.proportions2 = prepareProportionsAnalysisInstructor(measure.clean2);

plot(measure.proportions2$head.height.NA, measure.proportions2$height.NA, col = "blue",main = paste("Pl
```


Plot of head.height vs height of measure.proportions2 data
height.NA (cm)
100
120
140
160
180
15
20
25
30
35
head.height.NA (cm)

6.5.2 What is the mean of male or female proportions?

Here the newly created proportions of male and female get separated into new data frames, so that the mean and other statistical values of each can be viewed with the summary() command.

Findings: Mean of male and female proportions The findings are, that the mean proportions of head heights in the height of females in the data set is 158.3 cm. The mean value I get for males is 177.42 cm. When looking at the instructors data I get values of female and male respectively, 160.5 cm and 174.9 cm. So there is a little bit of a difference due to the data sets.

Below is the code for the summary statistics of male and female head height vs height proportions.

```
# Summary statistics of male and female proportions
meanMaleProportions(measure.proportions);
```

```
##    height.NA      head.height.NA    arm.reach          age
##  Min.   : 91.44   Min.   :16.51   Min.   :144.8   Min.   : 7.00
##  1st Qu.:169.75   1st Qu.:21.57   1st Qu.:211.0   1st Qu.:22.00
##  Median :177.00   Median :23.00   Median :221.6   Median :28.00
##  Mean   :177.42   Mean   :23.02   Mean   :223.3   Mean   :34.82
##  3rd Qu.:181.00   3rd Qu.:24.00   3rd Qu.:228.7   3rd Qu.:52.25
##  Max.   :464.52   Max.   :50.80   Max.   :587.1   Max.   :77.00
##     gender          proportions.head.and.height proportions.group
##  Length:72          Min.   :5.538               Length:72
##  Class :character   1st Qu.:7.150               Class :character
##  Mode  :character   Median :7.728               Mode  :character
##                     Mean   :7.734
##                     3rd Qu.:8.400
##                     Max.   :9.857
##   age.group
##  Length:72
##  Class :character
##  Mode  :character
##
##
##
```

```
meanFemaleProportions(measure.proportions);
```

```
##    height.NA      head.height.NA    arm.reach          age
##  Min.   : 36.0   Min.   : 6.50   Min.   : 57.0   Min.   : 5.00
##  1st Qu.:156.0   1st Qu.:20.00   1st Qu.:194.7   1st Qu.:21.00
##  Median :162.0   Median :21.00   Median :202.2   Median :28.50
##  Mean   :158.3   Mean   :21.15   Mean   :198.4   Mean   :36.19
##  3rd Qu.:167.7   3rd Qu.:22.06   3rd Qu.:209.8   3rd Qu.:53.00
##  Max.   :180.6   Max.   :31.00   Max.   :234.5   Max.   :94.00
##     gender          proportions.head.and.height proportions.group
##  Length:80          Min.   :5.138               Length:80
##  Class :character   1st Qu.:7.078               Class :character
##  Mode  :character   Median :7.600               Mode  :character
##                     Mean   :7.530
##                     3rd Qu.:8.140
##                     Max.   :9.278
##   age.group
##  Length:80
```

```
##  Class :character
##  Mode  :character
##
##
##
```

```
# same step for instructors data
meanMaleProportions(measure.proportions2);
```

```
##    height.NA     head.height.NA   arm.span.NA         age
##  Min.   :132.7   Min.   :17.78   Min.   : 65.0   Min.   : 7.0
##  1st Qu.:170.2   1st Qu.:21.59   1st Qu.:167.6   1st Qu.:22.0
##  Median :177.8   Median :23.00   Median :177.2   Median :26.0
##  Mean   :174.9   Mean   :22.96   Mean   :174.9   Mean   :32.6
##  3rd Qu.:181.9   3rd Qu.:24.00   3rd Qu.:183.4   3rd Qu.:44.5
##  Max.   :191.1   Max.   :38.10   Max.   :224.0   Max.   :77.0
##     gender          proportions.head.and.height proportions.group
##  Length:114         Min.   :4.467               Length:114
##  Class :character   1st Qu.:7.163               Class :character
##  Mode  :character   Median :7.685               Mode  :character
##                     Mean   :7.697
##                     3rd Qu.:8.325
##                     Max.   :9.857
##   age.group
##  Length:114
##  Class :character
##  Mode  :character
##
##
##
```

```
meanFemaleProportions(measure.proportions2);
```

```
##    height.NA     head.height.NA   arm.span.NA         age
##  Min.   : 93.0   Min.   :15.24   Min.   : 61.5   Min.   : 3.00
##  1st Qu.:156.0   1st Qu.:20.00   1st Qu.:154.5   1st Qu.:21.00
##  Median :162.3   Median :21.00   Median :161.8   Median :28.00
##  Mean   :160.5   Mean   :21.39   Mean   :160.5   Mean   :35.51
##  3rd Qu.:167.7   3rd Qu.:22.29   3rd Qu.:170.4   3rd Qu.:52.25
##  Max.   :181.0   Max.   :33.02   Max.   :198.5   Max.   :94.00
##     gender          proportions.head.and.height proportions.group
##  Length:108         Min.   :4.650               Length:108
##  Class :character   1st Qu.:7.053               Class :character
##  Mode  :character   Median :7.642               Mode  :character
##                     Mean   :7.585
##                     3rd Qu.:8.157
##                     Max.   :9.278
##   age.group
##  Length:108
##  Class :character
##  Mode  :character
##
##
##
```

6.5.3 Do group age and proportions match?

Create data frame with newly created 3 columns A new data frame will be created now with only the newly created columns.

Comparison age vs proportions group Here the analysis gets interesting. Now we compare if the two ways of assigning groups are matching in the assigned groups. The one way is assigning it by proportions of height/head.height and gender, the other way by age and gender. The result is a percentage of how many assigned labels match.

Findings: Comparison age vs proportions group The results give a percentage of my data give 90.13 %. This means that 9 out of 10 assignments using method 1 to assign by proportions and the method 2 to assign by age will give the same result. The professors data has a slightly better result on this analysis with 90.54 %.

Below the code for the comparison of group age and group proportions is shown.

```
# Do the 2 groups assigned match
match.percentage = matchingGroups(measure.proportions);
match.percentage;
```

```
## [1] 90.13158
```

```
# same step for instructors data
match.percentage2 = matchingGroups(measure.proportions2);
match.percentage2;
```

```
## [1] 90.54054
```

What percentage of people is considered 'heroic'?

Assign heroic based on head.height vs height proportions This analysis is based on the calculation of how many times the head height fits into the whole body height. Then the proportion values will be grouped into different groups, these are: Group 'F': There the person is most likely a female adult because the proportions are bigger than what kids usually have, which is above 6.5 and below or equal to 8.5. Group 'M': There the person is most likely a male adult because the proportions are bigger than what kids usually have, which is above 6.5 and below or equal to 8.5. Group 'K': There the person is most likely a child (male or female) because the proportions are smaller or equal to 6.5. Group 'H': There the person is most likely an adult because the proportions are heroic above 8.5. The result is a percentage of how many assignments from the data are 'heroic'.

Findings: Heroic The results give a percentage of about 14.47 %. This means that 14 out of every 100 people is considered heroic, and therefore taller than the average height when looking at the head.height vs height proportions only! The professor's data indicates a percentage of 16.96 %.

Below the code for the comparison of group age and group proportions is shown.

```
# What percentage of people is considered heroic?
h.percentage = heroicPeople(measure.proportions);
h.percentage;
```

```
## [1] 14.47368
```

```
# same step for instructors data
h.percentage2 = heroicPeople(measure.proportions2);
h.percentage2;
```

```
## [1] 16.96429
```

What percentage is male out of assigned heroic people? From the assigned percentage of heroic people over the total assignments, now the percentage of that assigned heroic people that are male will be found out with a function. And the assignments left have to be the female ones.

Findings: Heroic male percentage The results give a percentage of the people that are male out of the assigned heroic people. This is found to be 61.54 % (male out of heroic), with then being 38.46 % females heroic. These values are the same for the instructors data in this analysis.

Below the code of percentages of heroic males and heroic females are given.

```
# What percentage of heroic people are male
heroic.male.p = percentageOfHeroicPeopleBeingMale(measure.proportions);
heroic.female.p = 100 - heroic.male.p;
heroic.male.p;
```

```
## [1] 61.53846
```

```
heroic.female.p;
```

```
## [1] 38.46154
```

```
# same for instructors data
heroic.male.p2 = percentageOfHeroicPeopleBeingMale(measure.proportions2);
heroic.female.p2 = 100 - heroic.male.p2;
heroic.male.p2;
```

```
## [1] 61.53846
```

```
heroic.female.p2;
```

```
## [1] 38.46154
```

Assign heroic based on height Here a major part is the reference to a website "https://ourworldindata.org/human-height" where the comparison data for the assignment was taken from. There is says that the average height for male around the world is 171 cm and for female 159 cm. So every height in our data that is higher will be considered as heroic. Also specified is the gender to get a percentage of how many heroic male there are compared to female.

Findings: Heroic Height The results give a percentage of the people that are male out of the assigned heroic people. That percentage is 51.55 %. Then female heroic people compared to average height are 48.45 %. The percentage of the instructors data set varies a little bit as it has 55.33 % males and 44.67 % females categorized as heroic.

Below the code of percentages of heroic males and heroic females are given.

```
# Assign people heroic based on height and get percentage
heroic.m.p = assignHeroicByHeightMalePercentage(measure.proportions);
heroic.f.p = 100 - heroic.m.p;
heroic.m.p;
```

```
## [1] 51.54639
```

```
heroic.f.p;
```

```
## [1] 48.45361
```

```
# same for instructors data
heroic.m.p2 = assignHeroicByHeightMalePercentage(measure.proportions2);
heroic.f.p2 = 100 - heroic.m.p2;
heroic.m.p2;
```

```
## [1] 55.33333
```

```
heroic.f.p2;
```

```
## [1] 44.66667
```

6.5.4 How does arm span relate to height?

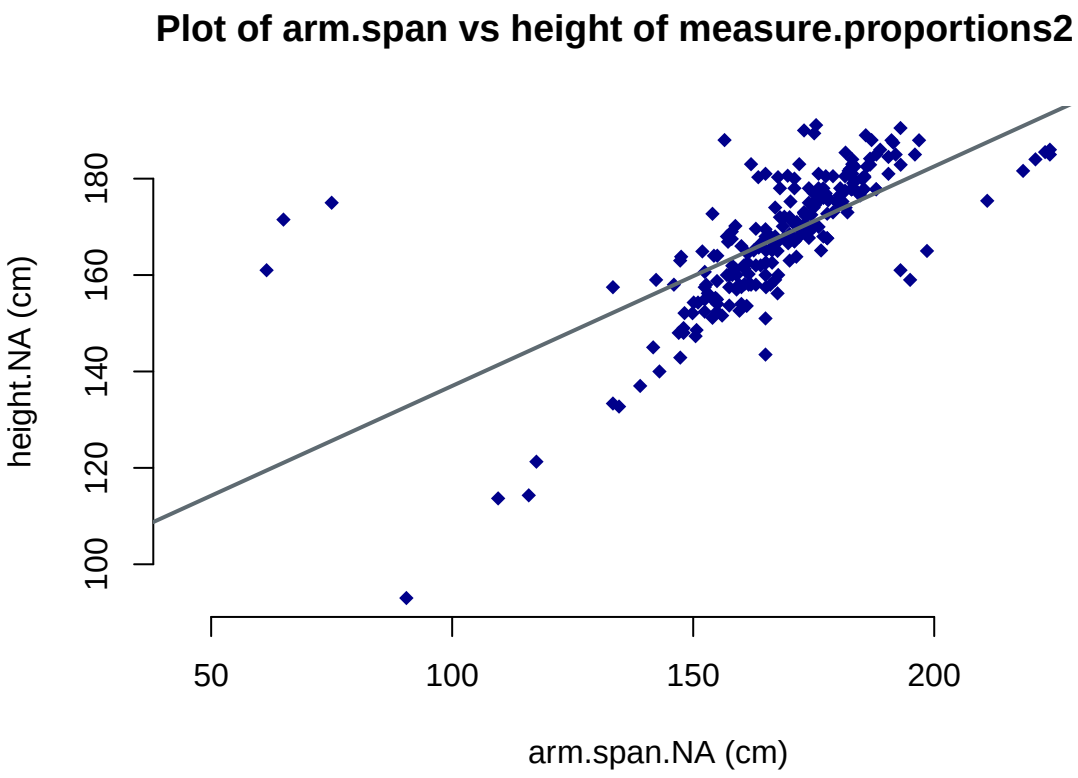
Another research question's interest is the relationship between arm.span and height.

Plotting variables This is a really interesting analysis. Since we know from the Figure 4. that a person should have around the same height and arm span it should be also verified in this sections analysis. A human proportions are given like this.

Findings of plotting variables When making a plot and viewing a summary statistics the plot we see is a linear line graph. It clearly shows a linear relationship between those two variables. The axis labels are also the same, which means that it is nearly a slope of 1, which would be perfectly linear. The graph is shown as the OneGraphic.**??** The summary statistics also proves that because the mean, 1st and 3rd Quartile values are very close. And even the match percentage >90% analysis gives a very positive result. In my cleaned data set 90.13 % of the height and arm span measurement values are within a difference of 0.1 (10%). And in the professor's data set they are 90.54 %.

```
# Is there a linear relationship between the two variables? -> Yes
# plot(measure.proportions$arm.span.NA, measure.proportions$height.NA, col = "#981e32", pch = 18, asp =
# abline(lm(measure.proportions$height.NA~measure.proportions$arm.span.NA), col="#5e6a71", lwd = 2);
# summary(measure.proportions);

# same step for instructors data
plot(measure.proportions2$arm.span.NA, measure.proportions2$height.NA, col = "darkblue",pch = 18, asp =
abline(lm(measure.proportions2$height.NA~measure.proportions2$arm.span.NA), col="#5e6a71", lwd = 2)
```



```
summary(measure.proportions2);
```

```
##    height.NA     head.height.NA   arm.span.NA         age
##  Min.   : 93.0   Min.   :15.24   Min.   : 61.5   Min.   : 3.00
##  1st Qu.:160.0   1st Qu.:20.50   1st Qu.:158.9   1st Qu.:22.00
##  Median :169.0   Median :22.00   Median :169.7   Median :27.00
##  Mean   :168.0   Mean   :22.16   Mean   :168.0   Mean   :34.31
##  3rd Qu.:177.8   3rd Qu.:23.50   3rd Qu.:179.8   3rd Qu.:50.25
##  Max.   :191.1   Max.   :38.10   Max.   :224.0   Max.   :94.00
##     gender          proportions.head.and.height proportions.group
##  Length:232         Min.   :4.467               Length:232
##  Class :character   1st Qu.:7.130               Class :character
##  Mode  :character   Median :7.685               Mode  :character
##                     Mean   :7.655
##                     3rd Qu.:8.234
##                     Max.   :9.857
##   age.group
##  Length:232
##  Class :character
##  Mode  :character
##
##
##
```

Below is the code for the comparison of a 90% match.

```
# # Is there a relationship between arm.span and height?
# match.percentage.90 = relateArmSpanAndHeight(measure.proportions);
#
# # same step for instructors data
# match.percentage.90.2 = relateArmSpanAndHeight(measure.proportions2);
```

6.5.5 Proportions of all measurements vs head.height

This is of additional interest for additional observations.

Below the code for the comparison of group age and group proportions is shown.

```
# Proportions of all measurements vs head.height
measure.head.proportions = proportionsAllColumnsVSHeadHeight(measure.clean);
# measure.head.proportions;

# same step for instructors data
# measure.head.proportions2 = proportionsAllColumnsVSHeadHeight(measure.clean2);
```

6.5.6 Statistical analysis

Correlation and t-test This section includes a correlation test and t-test upon some variables of the data. Those variables are the main variables of interest. Each correlation and t-test is done on 2 variables only.

Findings: Correlation and t-test My data: The correlation test between height and head height gives a correlation score of 0.6565967 with a really small p-value. Therefore, these 2 variables have a quite strong correlation. When looking at height and age, I would expect only a small correlation, because kids that are young are not grown up and relatively short. When grown up the difference of height is high and usually when getting really old then the height does not increase anymore, it rather decreases. So that is not a linear relationship at all. This is true, because the correlation score is only 0.05830303 at a p-value of 0.417. Whereas height and arm span do have a strong correlation again. The height does almost exactly match the arm span. Therefore the score is 0.8785036 and the p-value again very small, telling the correlation is strong. When looking at the t-test, then we see again, that the variables height and head height, height and age are different from each other when looking at the sample means, whereas the height and arm span do not differ much. Then have almost the same sample means and a high p-value instead.

Below the code for the correlation test and t-test is shown.

```
# # correlation
# cor.test(measure.proportions$height.NA, measure.proportions$head.height.NA);
# cor.test(measure.proportions$height.NA, measure.proportions$age);
# cor.test(measure.proportions$height.NA, measure.proportions$arm.span.NA);
#
# # same for instructor data
# cor.test(measure.proportions2$height.NA, measure.proportions2$head.height.NA);
# cor.test(measure.proportions2$height.NA, measure.proportions2$age);
# cor.test(measure.proportions2$height.NA, measure.proportions2$arm.span.NA);


# # T-test
# t.test(measure.proportions$height.NA, measure.proportions$head.height.NA);
# t.test(measure.proportions$height.NA, measure.proportions$age);
# t.test(measure.proportions$height.NA, measure.proportions$arm.span.NA);
#
# # same for instructor data
# t.test(measure.proportions2$height.NA, measure.proportions2$head.height.NA);
# t.test(measure.proportions2$height.NA, measure.proportions2$age);
# t.test(measure.proportions2$height.NA, measure.proportions2$arm.span.NA);
```

Correlation table Below is the code to generate the correlation table that you see in Section 5.

```
library(devtools);       # required for source_url
path.humanVerseWSU = "https://raw.githubusercontent.com/MonteShaffer/humanVerseWSU/"
source_url( paste0(path.humanVerseWSU,"master/misc/functions-project-measure.R") );
path.project = "C:/_git_/WSU_STATS419_FALL2020/Project1/michaela_project_measure/project-measure/";
path.tables = paste0(path.project,"tables/");
  createDirRecursive(path.tables);


file.correlation = paste0(path.tables,"measure-correlation-table.tex");

outfile = "C:/_git_/SECRETS/measure-students.rds";
measure.clean = readRDS(outfile);
```

```
myData = as.matrix(measure.clean[,c(4,12,5)]);
# height, arm.span.NA, head.height.NA
# https://www.overleaf.com/read/srzhrcryjpwm
# keepaspectratio of include graphics
# could scale \input if still too big ...
# https://tex.stackexchange.com/questions/13460/scalebox-knowing-how-much-it-scales#13487
buildLatexCorrelationTable(myData,
  rotateTable = FALSE,
  width.table = 0.90, # bet for given data ... 0.95 when rotateTable = FALSE
                      # 0.60 when rotateTable = TRUE
  myFile = file.correlation,
  myNames = c("Height (cm)", "Arm Span (cm)", "Head Height (cm)"),
  myCaption = "Descriptive Statistics and Correlation Analysis");
Sys.sleep(2); # in case Knit-PDF doesn't like that I just created the file...
```

ENDNOTES

REFERENCES

Thomas, Diana M, David Galbreath, Maura Boucher, Krista Watts. 2020. Revisiting Leonardo da Vinci's Vitruvian Man Using Contemporary Measurements. *JAMA* **323**(22), 2342–2343.

TABLE OF CONTENTS