Fatma Baytar Data Analysis

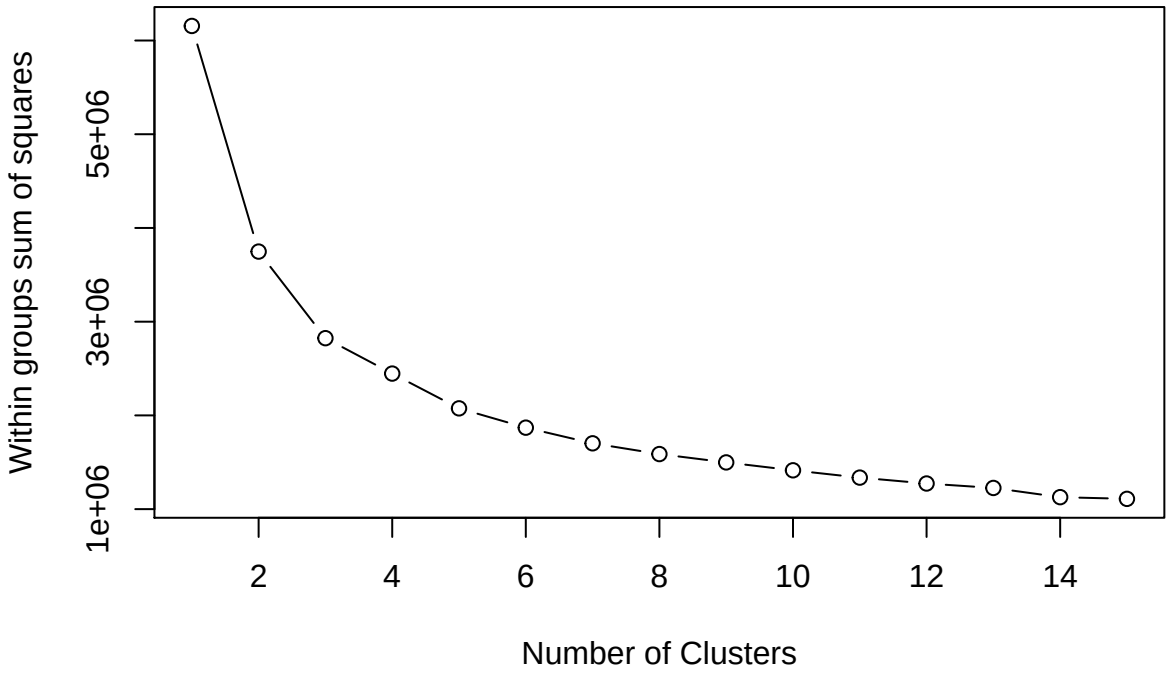
Jennifer Russell (jjr265) & Toshihiro Tokuyama (tt426)

Introduction

In clothing patternmaking, knowing the customers measurements can make fitting much more accurate, therefore ensuring a better fitting garmet. Most fashion brands currently operate under the assumption that bodies fit a standard model, when in reality body shapes come in many different shapes and sizes. When measuring for pants, the crotch curve measurement is not taken for privacy's sake, although this measurement can drastically impact the overall fit of the garment. To better design womens pants, we look at the measurements, including crotch curves, of 404 White woman from the Civilian American and European Surface Anthropomentry Resource (CAESAR) 3D body scan database to try to gain insight into the features of the crotch curves and how this knowledge may help design more accurate shapes to utilize in pattern making. We wanted to focus on predicted curves for a few groups of women since most brands use standard sizing and don't design custom pieces for their customers.

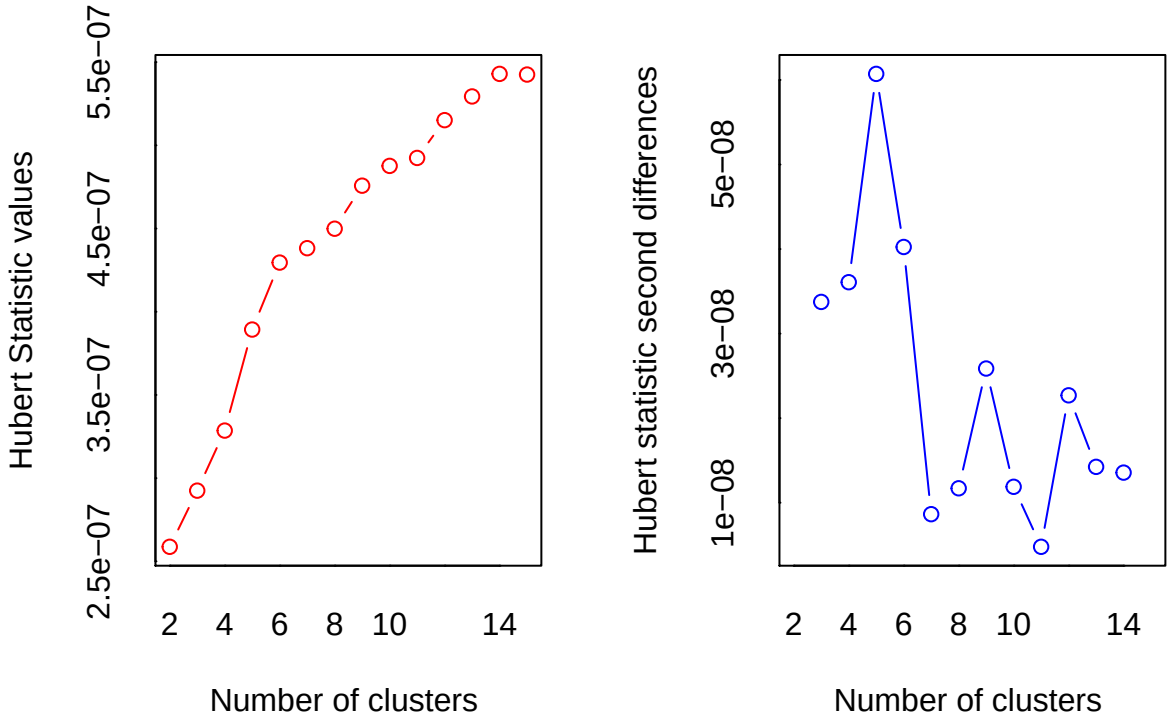
Forming Clusters

Inititally, in order to draw conclusions for women of different body shapes, we wanted to cluster the women based on their basic measurements: BMI, Height, and Max Hip. We also used the basic crotch measurements given in order to account for the fact that women of similar basic measurements like BMI and height can have very different body types.

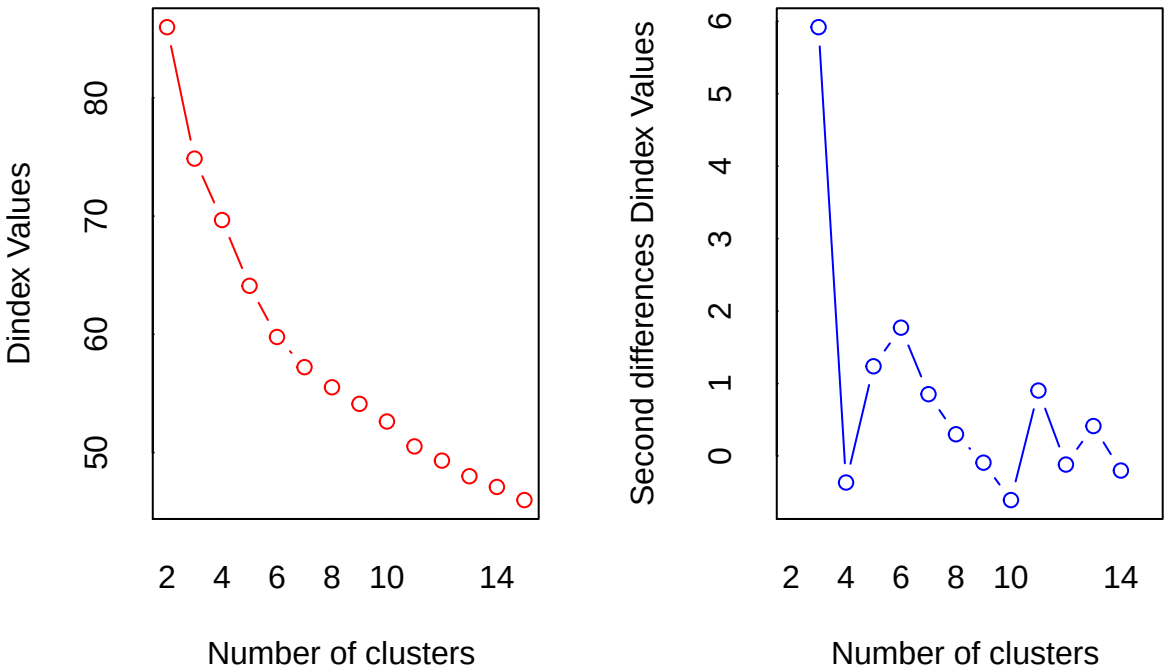


This graph shows us that after about 3 groups, there is little significant decrease in the within groups sum of squares. This suggests that 3 clusters might have a good fit.

We then use the library NBClust, which uses many indices to provide a majority-rule opinion about how many clusters to use based off the results of all the indices tests.

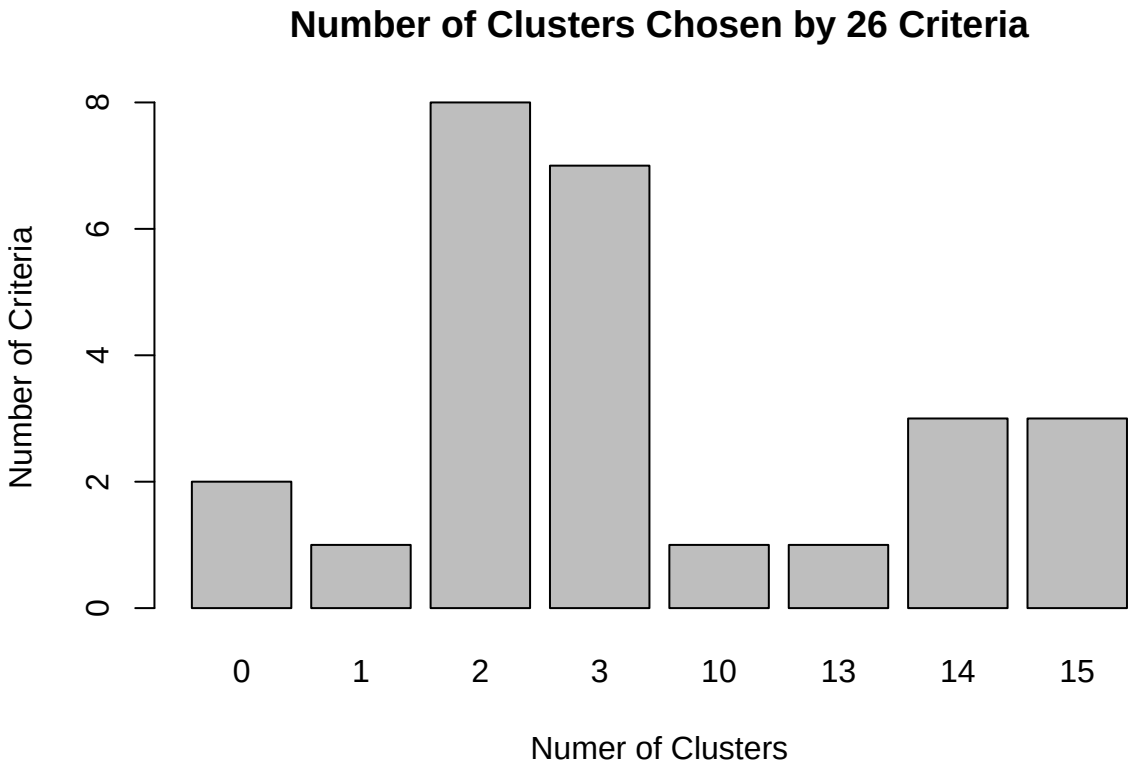



```
## *** : The Hubert index is a graphical method of determining the number of clusters.
##                 In the plot of Hubert index, we seek a significant knee that corresponds to a
##                 significant increase of the value of the measure i.e the significant peak in Hubert
##                 index second differences plot.
##
```




```
## *** : The D index is a graphical method of determining the number of clusters.
```

```
##                 In the plot of D index, we seek a significant knee (the significant peak in Dindex
##                 second differences plot) that corresponds to a significant increase of the value of
##                 the measure.
##
## *****************************************************************
## * Among all indices:
## * 8 proposed 2 as the best number of clusters
## * 7 proposed 3 as the best number of clusters
## * 1 proposed 10 as the best number of clusters
## * 1 proposed 13 as the best number of clusters
## * 3 proposed 14 as the best number of clusters
## * 3 proposed 15 as the best number of clusters
##
##                    ***** Conclusion *****
##
## * According to the majority rule, the best number of clusters is  2
##
##
## *****************************************************************
```




```
##
##  0  1  2  3 10 13 14 15
##  2  1  8  7  1  1  3  3
```

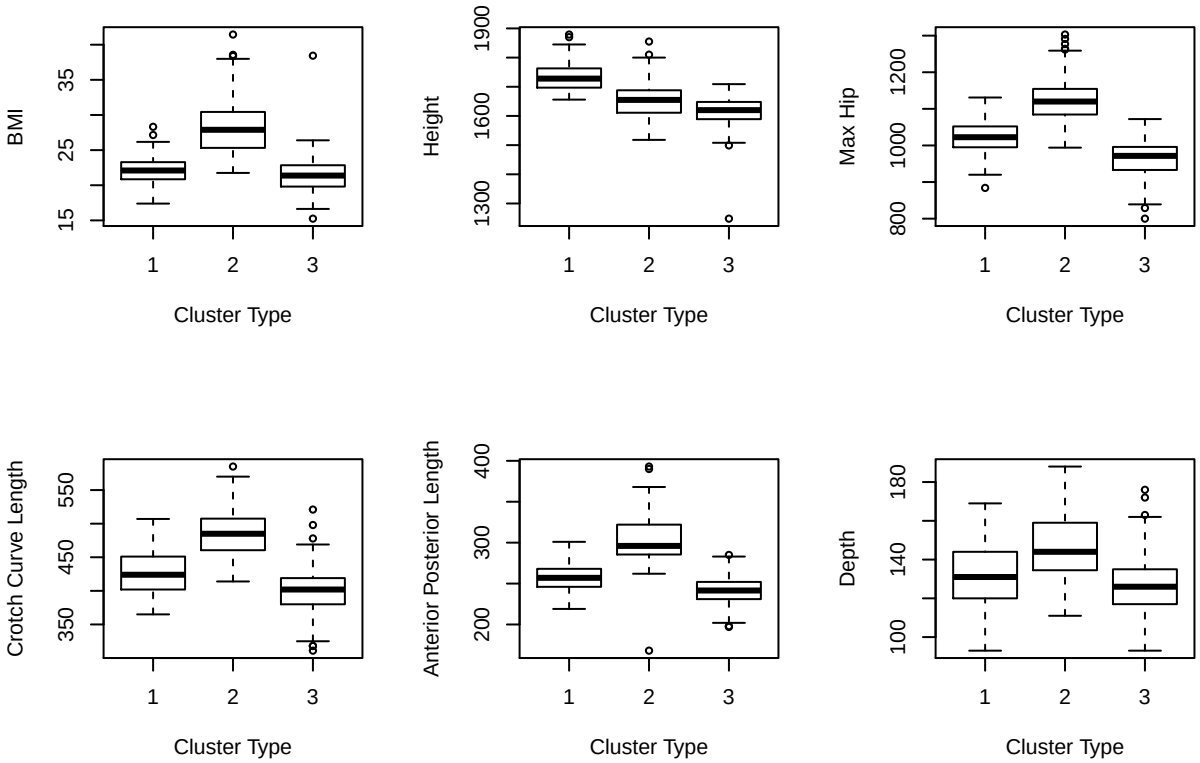
Since we arrived to the conclusion of 3 clusters using the NBClust package, we will stick with that in our further analysis.

K-Means Clusters

Next, we use the kmeans function in R to perform the clustering. We pass 3 as the argument for how many clusters since that was the conclusion of our prior analysis. To get an idea of the three clusters, shown below is the mean measurements for women in all of the three groups. Based off of this, we now have clusters of 3 clusters of women with similar measurements, and also what they're average crotch curve measurements would be. While the clusters did utilize measurements such as crotch length and crotch depth, we believe it would be feasible to determine with relatively high accuracy which cluster a future women would belong to off of her basic measurements.

```
##   Group.1      BMI   Height   Max Hip Crotch curve length at trochanter
## 1       1 22.06222 1732.008 1024.3607                           425.3443
## 2       2 28.34493 1654.956 1128.4176                           485.0549
## 3       3 21.49056 1616.495  965.9368                           400.6579
##   Anterior-posterior Length    Depth
## 1                  257.1721 132.4836
## 2                  304.0330 146.2747
## 3                  241.4632 126.2684
```

To get a better understanding of the three clusters created, we look at box-plots detailing the averages and range of measurements by women in the three clusters.



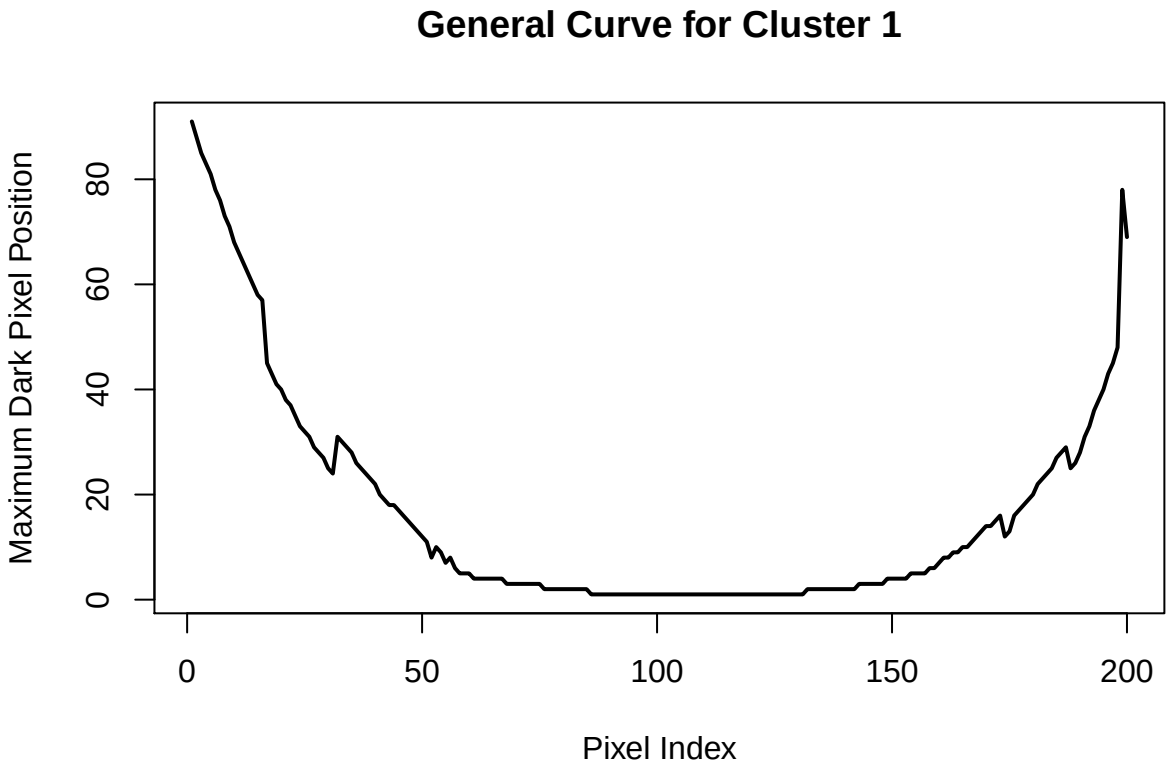
From the box plots we see a general trend: cluster 2 have the highest average values for all variables except height. However, making predictions from the box plots will be difficult because for most variables because the interquartile overlaps. If we were take make predictions of the cluster fits, using the max hip variable will provide the most accurate predictions as the interquartile range do not overlap.

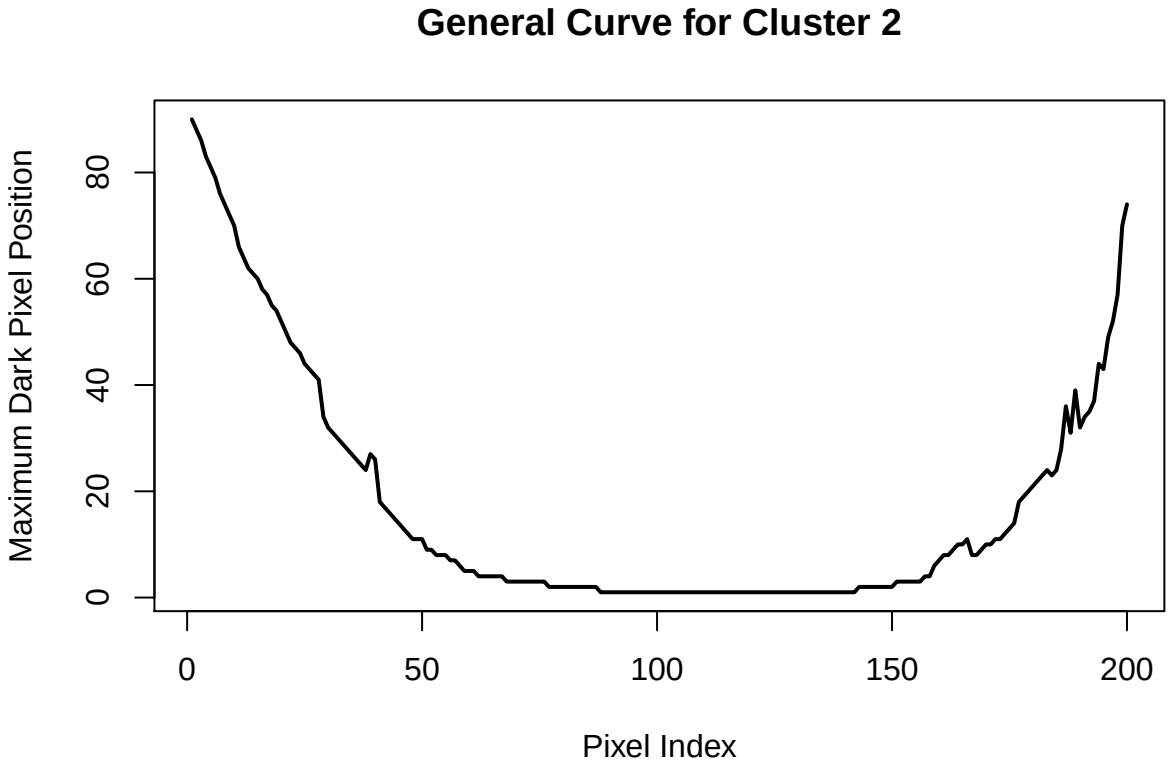
Creating a General Curve for Each Cluster

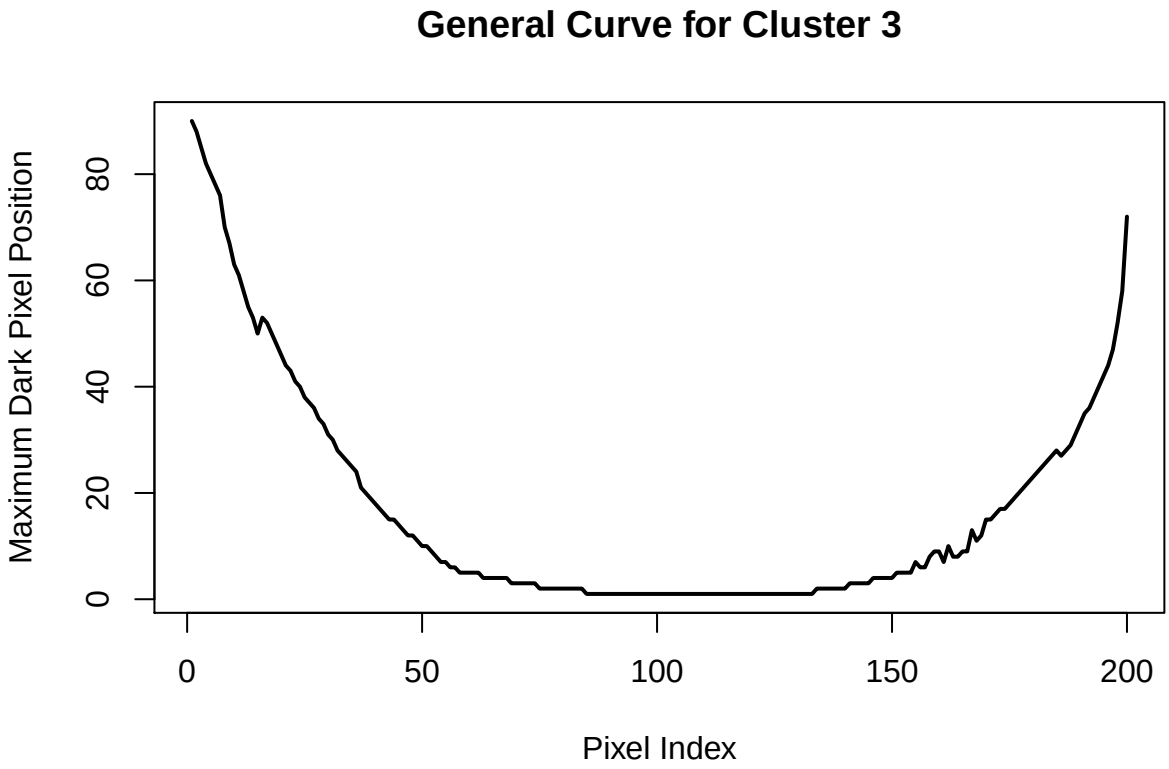
Now that we know the characteristics that seperates clusters, we try to plot a general curve shapes for each cluster. We first find the images that corresponds to each cluster fit by going through the image folder and searching the image that matches the observation.

Then to get an general curve shape, we will compute the average grey scale value for each pixel for each cluster fit. However, since all the image files have different pixels, we resize all images to 91x200 pixels. This process will lose some information for the images, possibly making the general curve shape less accurate.

We create a basic curve from the average grey scale matrix computed for each cluster fit by using maximum dark pixel position as value of y and x-value as the pixel index. These curves, along with the average measurements of women from the three groups, can be used to determine a more presice crotch curve measurements for large, but distinct, groups of women. The code used to create the curve images is taken from *Processing png image files into Curves* R notebook.







#

Characteristics of Each Curve

From the three plots we see that the difference of the curves is visible in the range of 25 to 40 and 160 to 190 for pixel index. The characeristics of each curve are described below:

Curve for Cluster 1

- There is a large bump aroud pixel 30.
- Two small bumps around pixel 175 and 190.

Curve for Cluster 2

- A small bump around pixel 40.
- Multiple bumps from pixel 160 to 190.

Curve for Cluster 3

- Small bumps around pixel 20
- Multiple small bumps from pixel 160 to 175.
- Compared to cluster 1 and cluster 2, the curve is generally smoothe.

Therefore, when drafting patterns the designers should pay close attention to these bumps will make the curve irregular.

Conclusion

In all, we recommend that patternmaking for womens pants focus on three main bodytype clusters. While each groups average curves are not perfect, they are more precise frameworks for crotch curve measurements. This additional accuracy in crotch curve measurements will significantly affect the fit of womens pants so pants can be more tailored to the wearer. We hope the identification of these three body type clusters and their respective crotch curves will help to inform future patternmaking and research regarding pant fit.