Shirt Sales Data: Exploratory Data Analysis

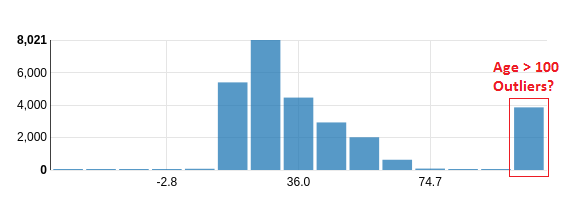
# Phase I: Descriptive Analysis

## Consumers Segments & Clusters

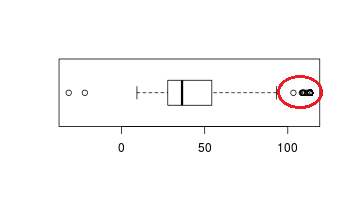
### Consumers segment by Age:

Histogram of Age:

The age may have outliers as marked in red.

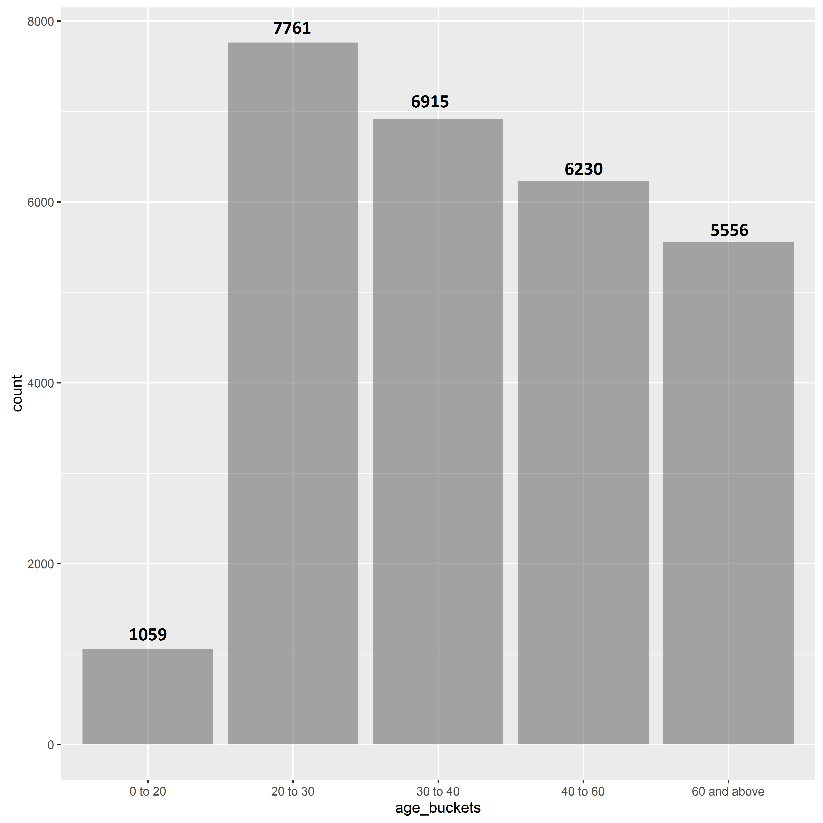


Box plot of Age:



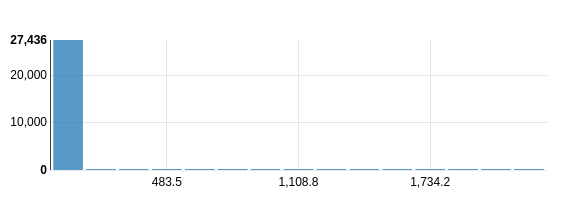
Binning Age into buckets:

Identified 5 user segments based on age:

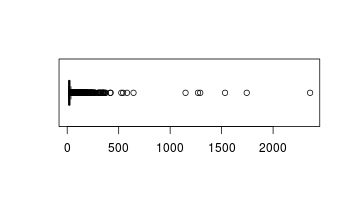


### Consumers segment by Expenditure:

Histogram of user expenditures (looks as if outliers present):

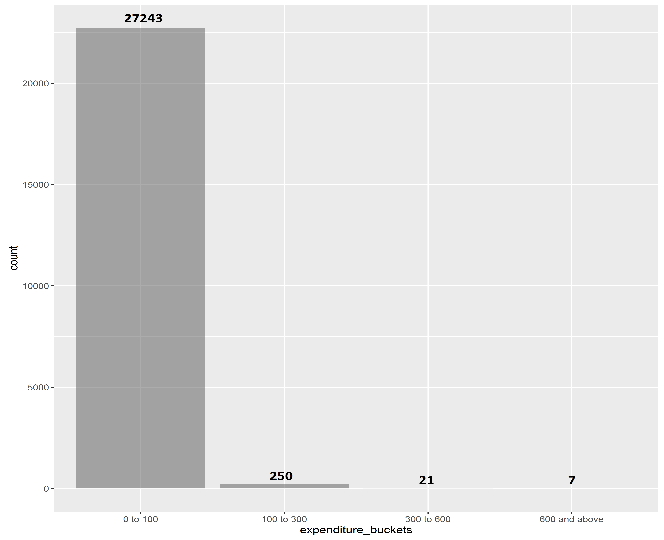


Box plot of User Expenditures:



Binning Expenditure into buckets:

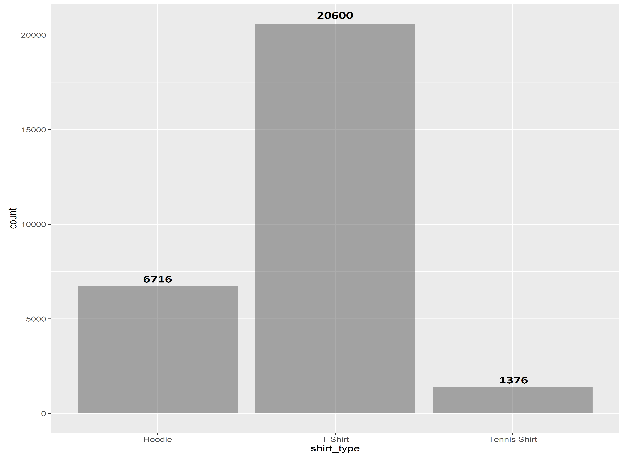
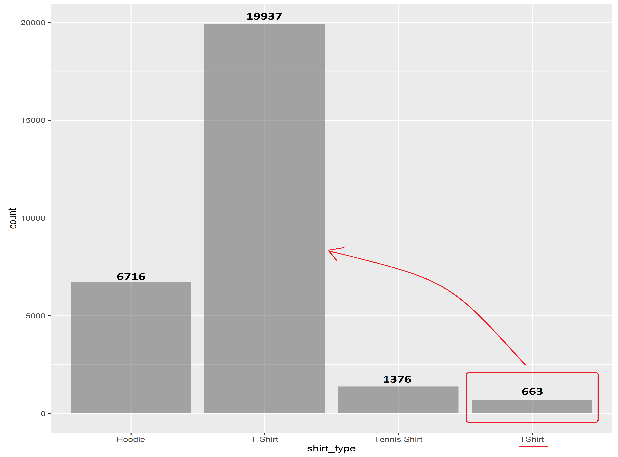
Identified 4 user segments based on expenditure:



## Products Segments & Clusters

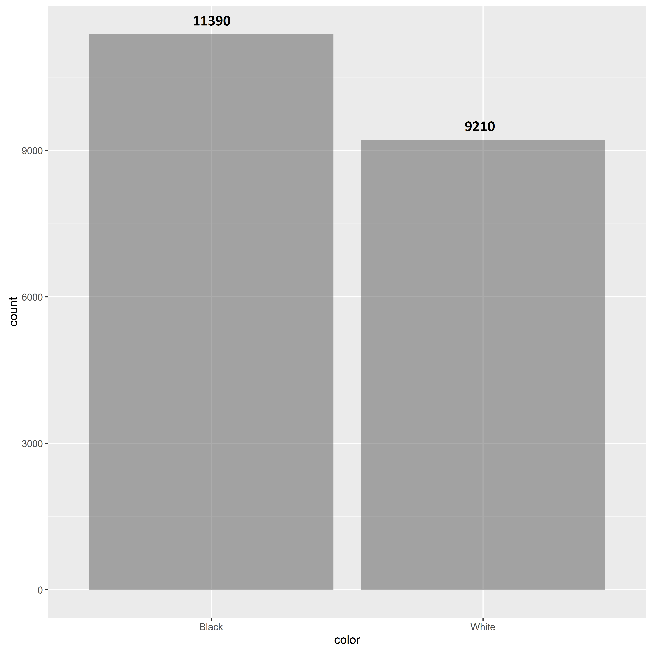
### Products segment by Shirt-type:

On clustering products by shirt-type, it is seen that there are some outlier categories (TShirt) that need to be merged with dominant category (T-Shirt):

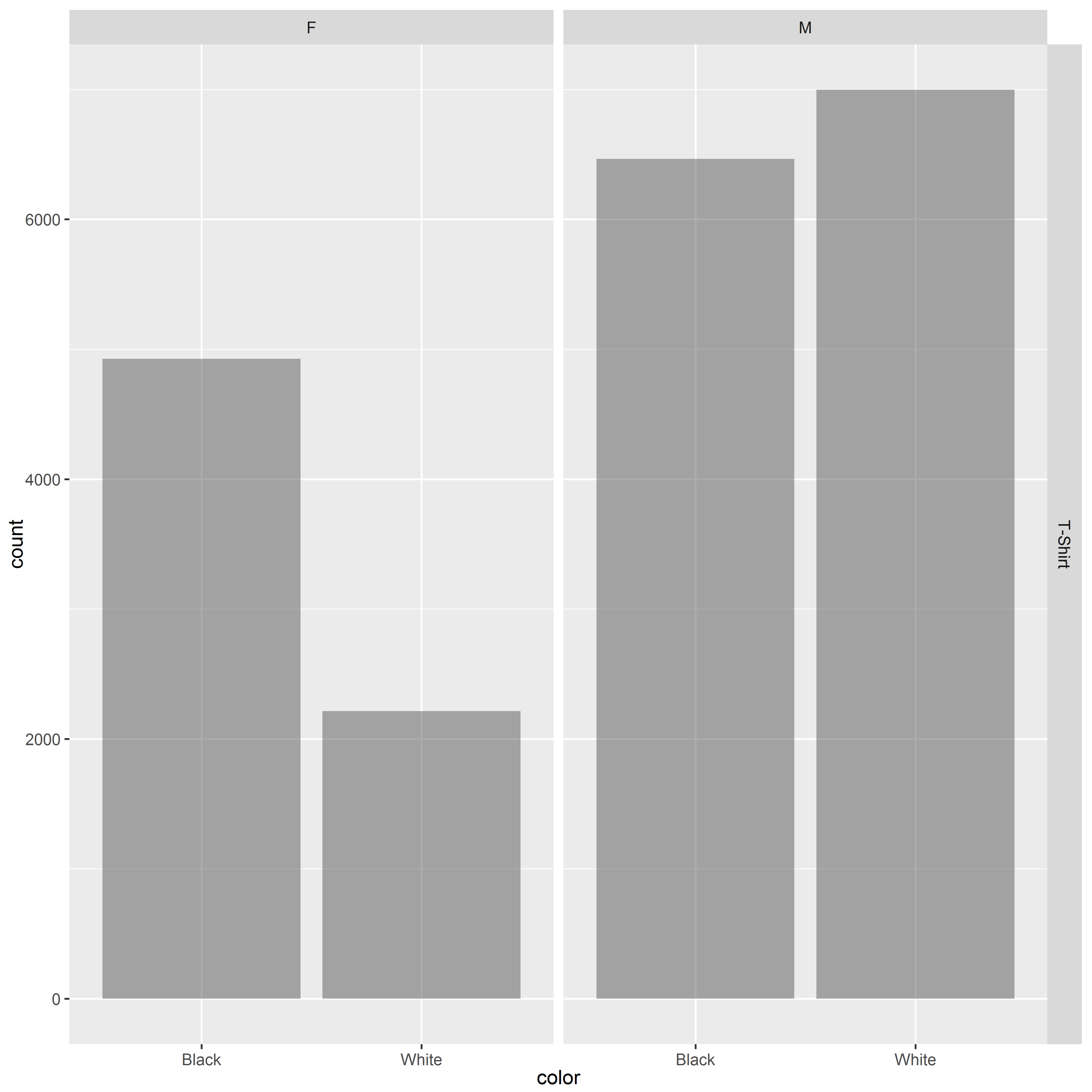


### Products segment by Shirt-colour:

Shirt colour segments after cleaning outliers:



Bar plot of shirt-color grouped by shirt-type and shirt-gender:



Inference drawn:

* Shirt colors available only for T-Shirts, not for Hoodies or Tennis Shirts
* Black T-Shirts is more popular with females
* White T-Shirts more popular with males
* Males T-Shirts sales numbers exceeds females sales by a good margin

# Study trends with time:

