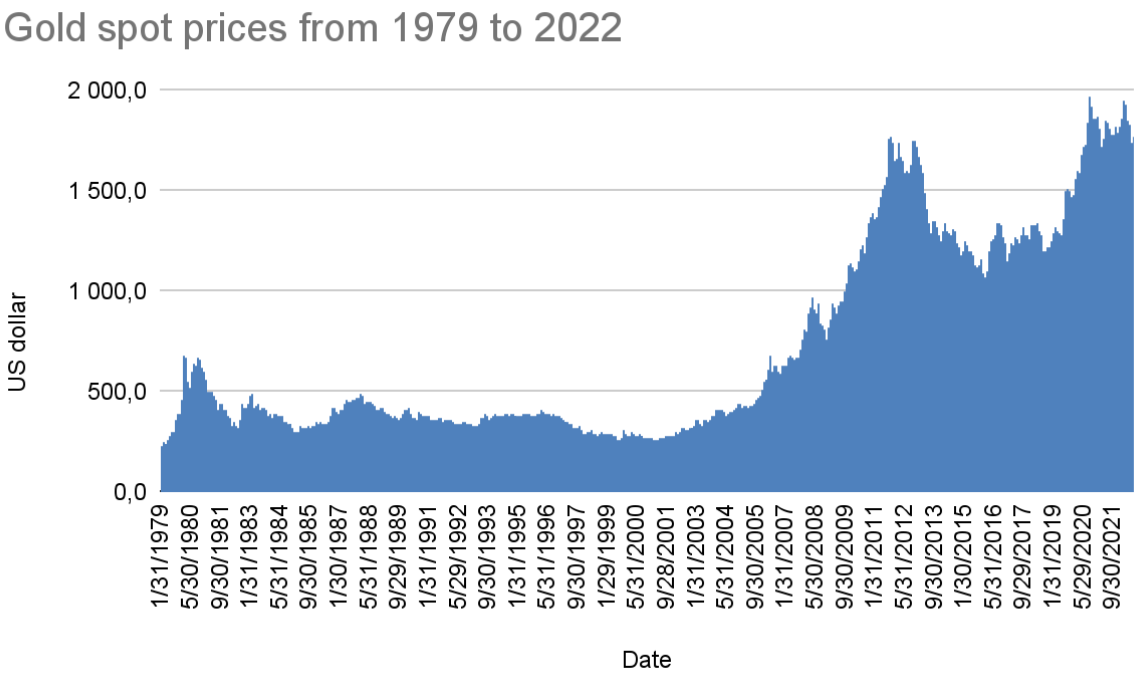
PROBABILITY THEORY & STATISTICS
Homework 1 – Descriptive Statistics
Hakob Hakobyan

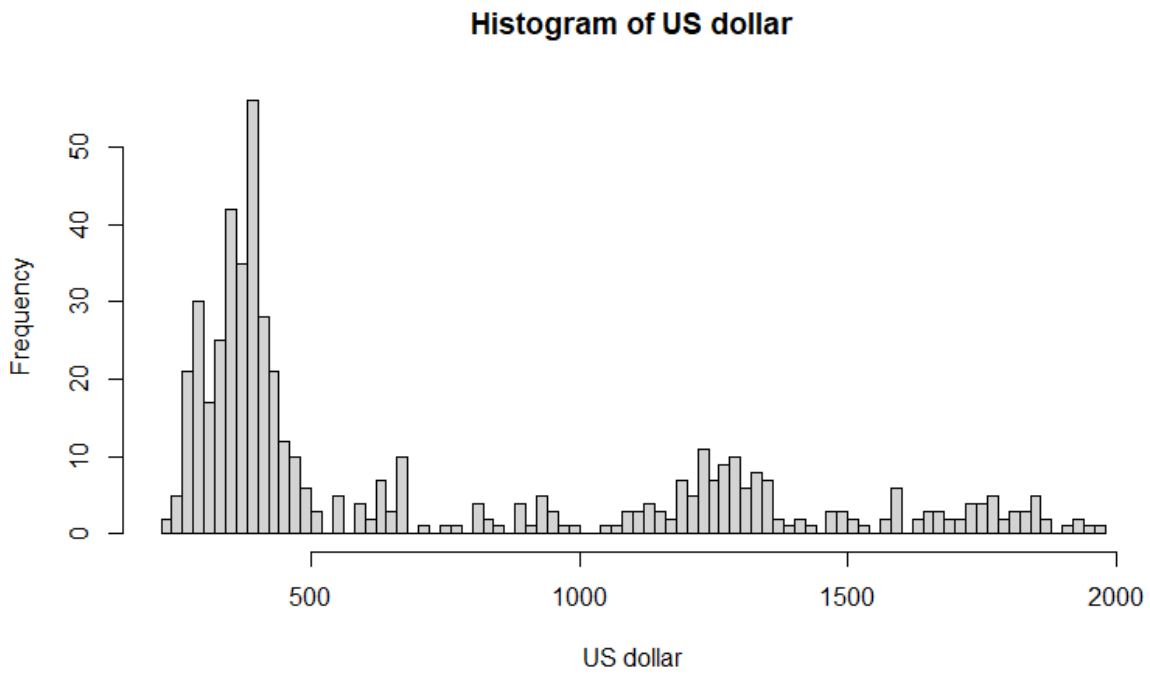
GOLD SPOT PRICES

The basic data set provides the gold price over a range of timeframes (daily, weekly, monthly, annually) going back to 1979, and in the major trading, producer, and consumer currencies. This data is taken from the web page of World Gold Council. I have then modified it removing data that we are not going to observe and leaving only monthly average gold prices expressed by different currencies. (You can find a similar data set for gold prices on Yahoo Finance web page, but the latter does not allow downloading.)

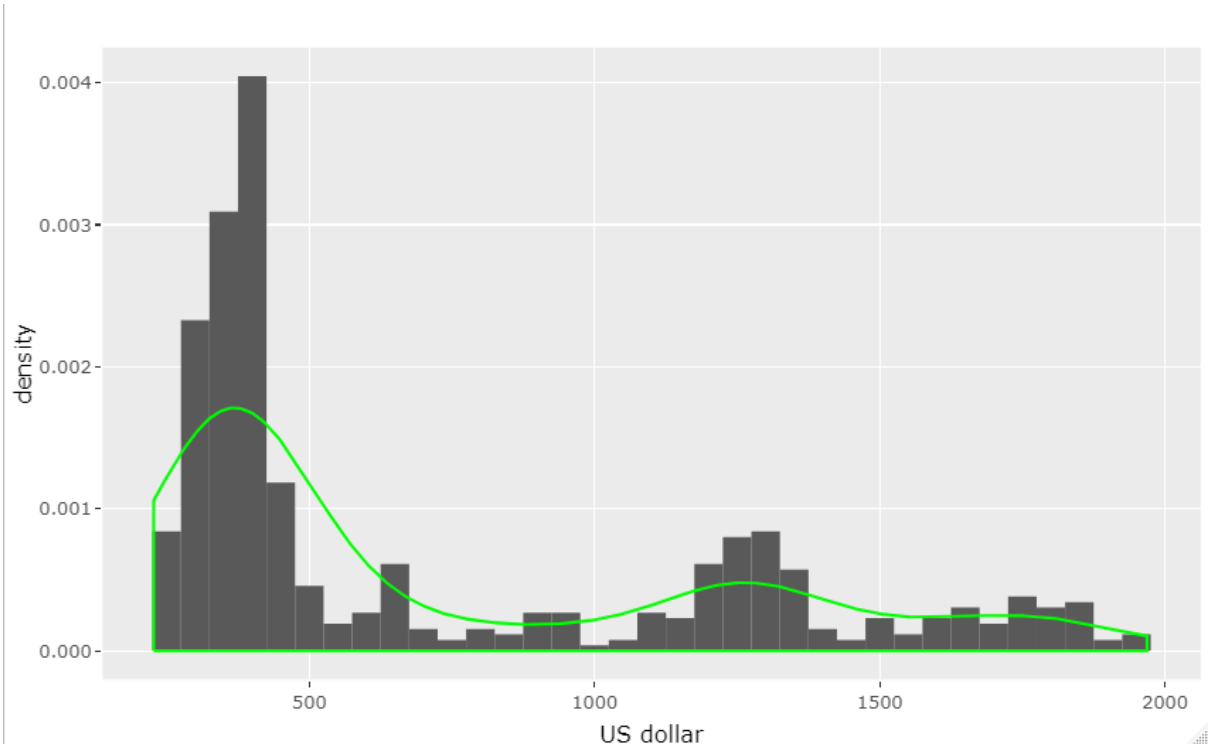
There are **524 observations** from 1979 to 2022 representing monthly averages of gold price for each year. And the monthly averages themselves are obtained by weekly and daily averages, which means this data set may be considered a sample for a larger population.



As you can see in the graph the price of gold has shown a sharp upward trend over the past 20 years, that's why I have observed the graphical data in two ways: one in general (1979-2022), and another one for the last 20 years (2002-2022).



This histogram shows us that there is a bigger concentration of values around the median ($420) not the mean. This is probably because the price of gold did not change dramatically during most of the period we observe, but only in the last 20 years.



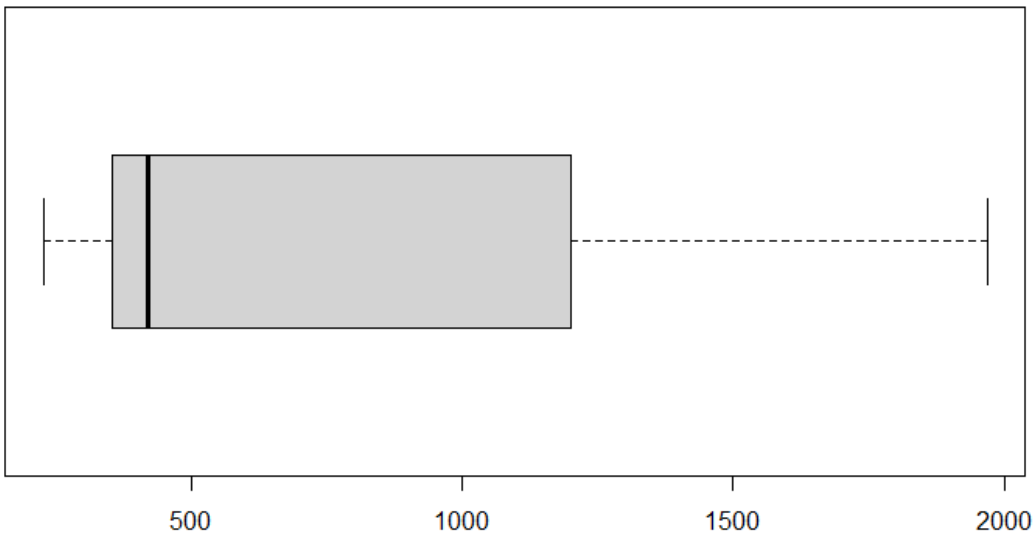
This is the same graph showing density of the set instead of frequency. We have got a multimodal distribution here.

Min value for gold prices over 43 years is **227.27** USD which was recorded on 1979-01-31. **Max** value is **1968.56** USD which was recorded on 2020-08-31.

Sample **mean** is equal to **721.03** USD, the **median** is **420.31** USD and the **mode** is **384,13** USD. Trimmed mean is 544 USD (when trim = 0.25). This kind of layout for mean, median and mode shows a positive skew. We can confirm this by looking at the graphs. Indeed when we calculate it we get a **positive skewness** value equal to **0.97**. Kurtosis is 2.5.

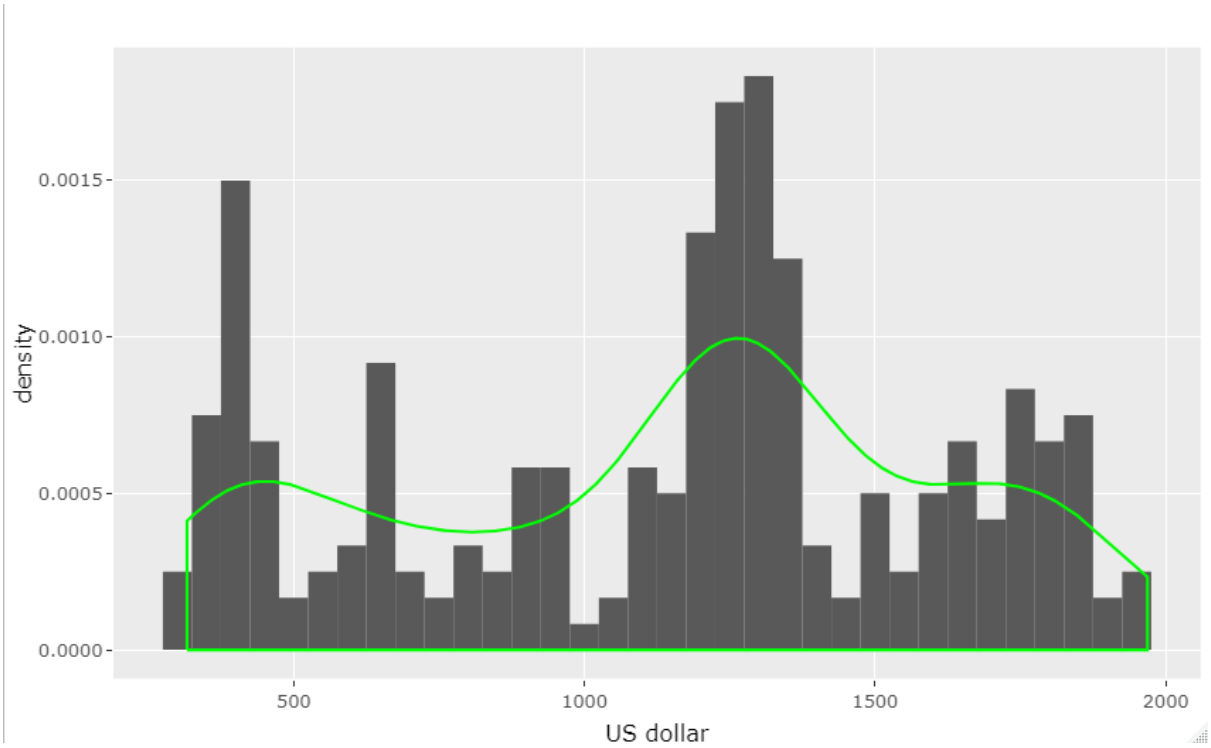
Standard deviation is 499.78 USD.

1st Qu. = 355, 3rd Qu. = 1200



Because of the wide range from the 1st and the 3rd quartiles (or a wide IQR), we don't have any outliers in our data set. It seems logical as we know that we are observing monthly average data. We can assume that if we observe data sets on a daily basis we will probably have more outliers.

Related to the issue mentioned above, finally let's take a look at the histogram for the last 20 years.



As we can see we have a multimodal distribution here too. In this case it seems like we have got a slightly negative skew, that is higher values are more than lower values. Anyway this graph seems better to observe than the previous one, because in recent decades the use of gold is also different, so the higher demand raises the price of gold in the market, and this is normal, just as the low price was normal for 1980-1990s.