Analysis of GDP growth of Asia

Han Zhang

Data Set Overview

The data set was gotten from Kaggle. It is a summation of the GDP activity of different countries in Asia It is further divided into subregions and years. The data set contains information from the year 2016 to 2020, as well as predicted GDP information for years 2021 and 2022 based on information at that time.

Data Prep

First the csv file was found on kaggle.com. The csv file was downloaded and put in the same directory as the working R file, which allowed the R file to find and read the csv file. There were a few modifications made to the file so that the data would be more accessible and readable. For example some categories were renamed to facilitate ease of access, and pieces of blank information were omitted so the data would be more accurate. After making the entire data set more accessible subsets and data wrangling methods were applied so that only relevant information was accessed for each catagory of analysis.

GDP break down of Asia GDP year by year

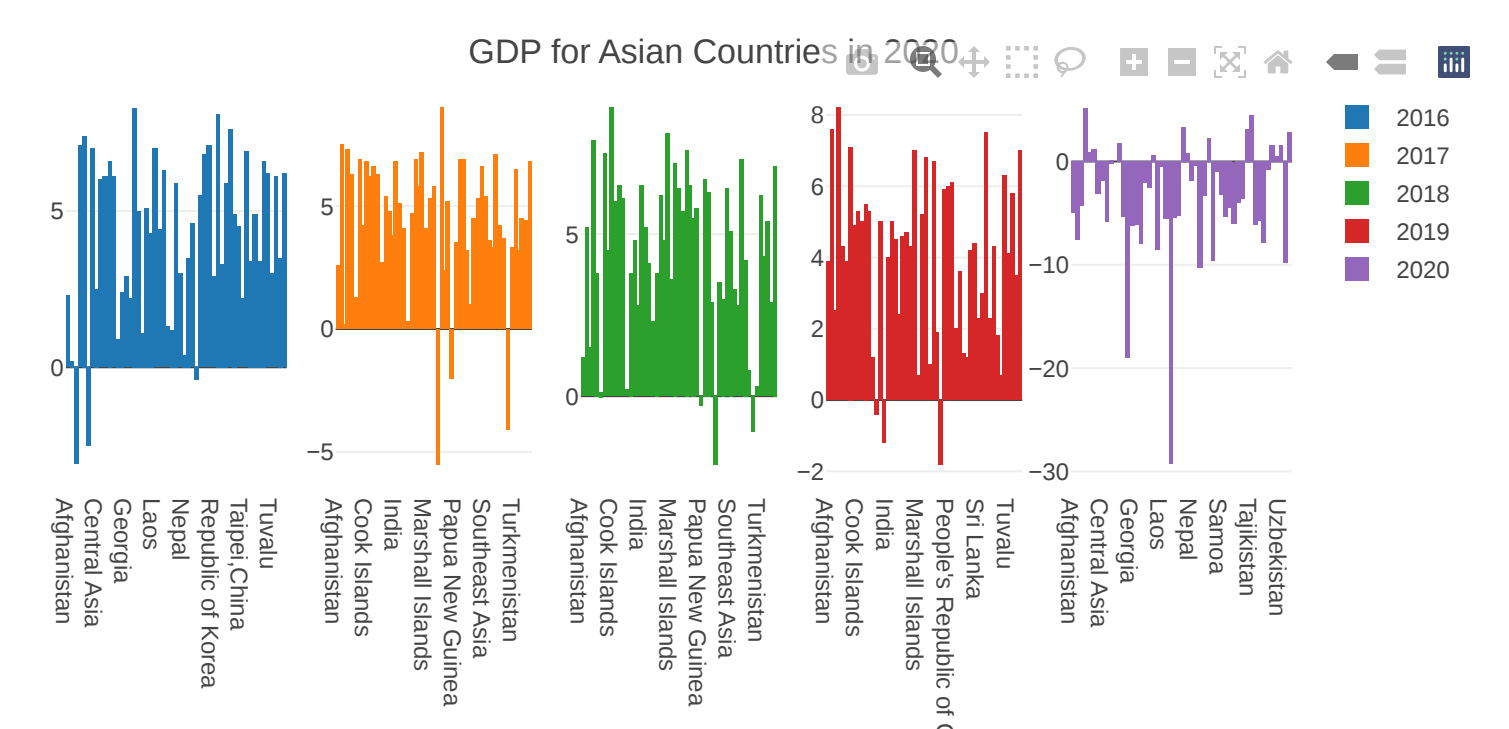
The first plot below shows the different GDP of various countries in Asia for the year 2016. The country with the highest GDP growth is India with 8.3% and the lowest is Armenia with a growth of 0.2%, while the countries with a negative growth are Azerbaijan at -3.1%, Brunei at -2.5%, and Palau at -0.4%

The second plot shows the same information for year 2017. In this plot we can see that the country with the highest GDP is Nepal at 9%, and the lowest Kiribati at .3%, while Nauru, Palau, and Timor-Leste have growths in the negatives

The third plot shows the information for year 2018, The country with the highest GDP is the Cook Islands at 8.9%, the lowest is Brunei at .1%, while the negative countries are Samoa, and Timor-Leste.

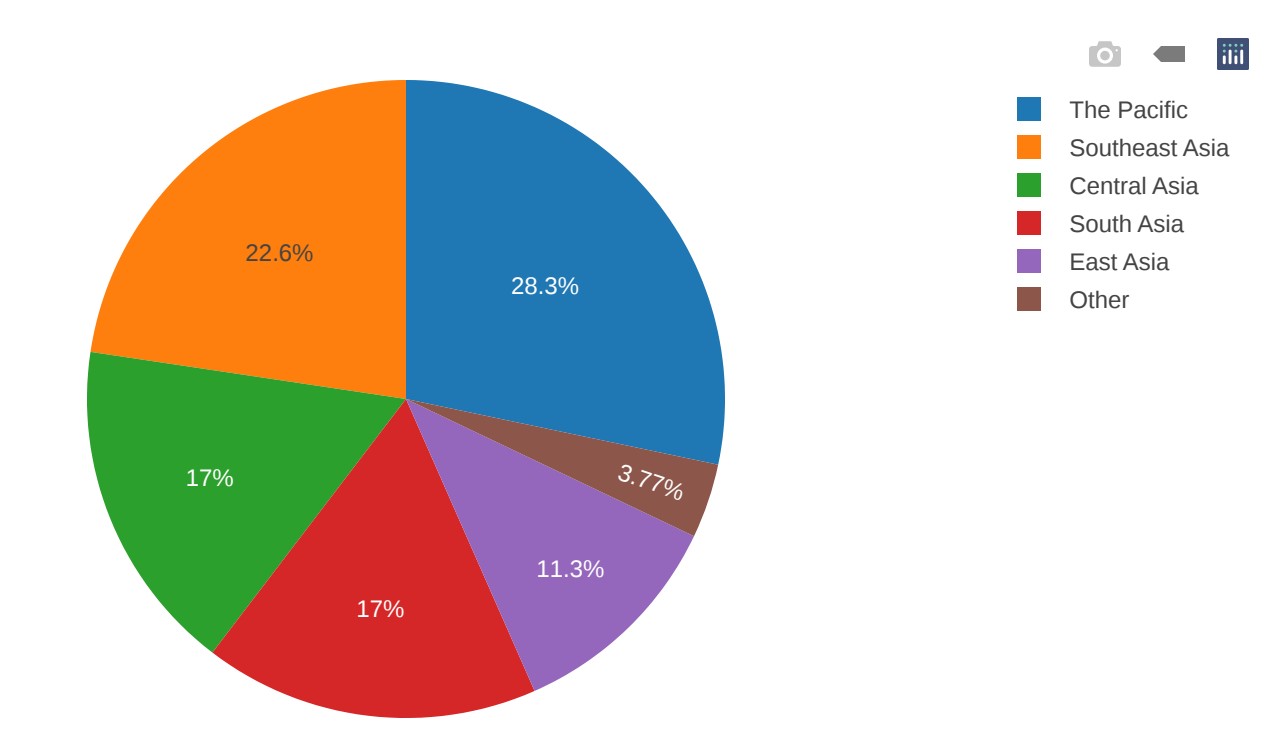
Continuing on the fourth plot shows the same information for year 2019. The highest GDP again is Bangladesh and the lowest are the Marshall Islands and Tonga at .7%. The negative countries are Fiji, Hong Kong, and Palau.

The last plot shows information for year 2020 and it can be seen that a majority of countries here have negative growth with few countries in the low positives.



Divison of countries by subregion

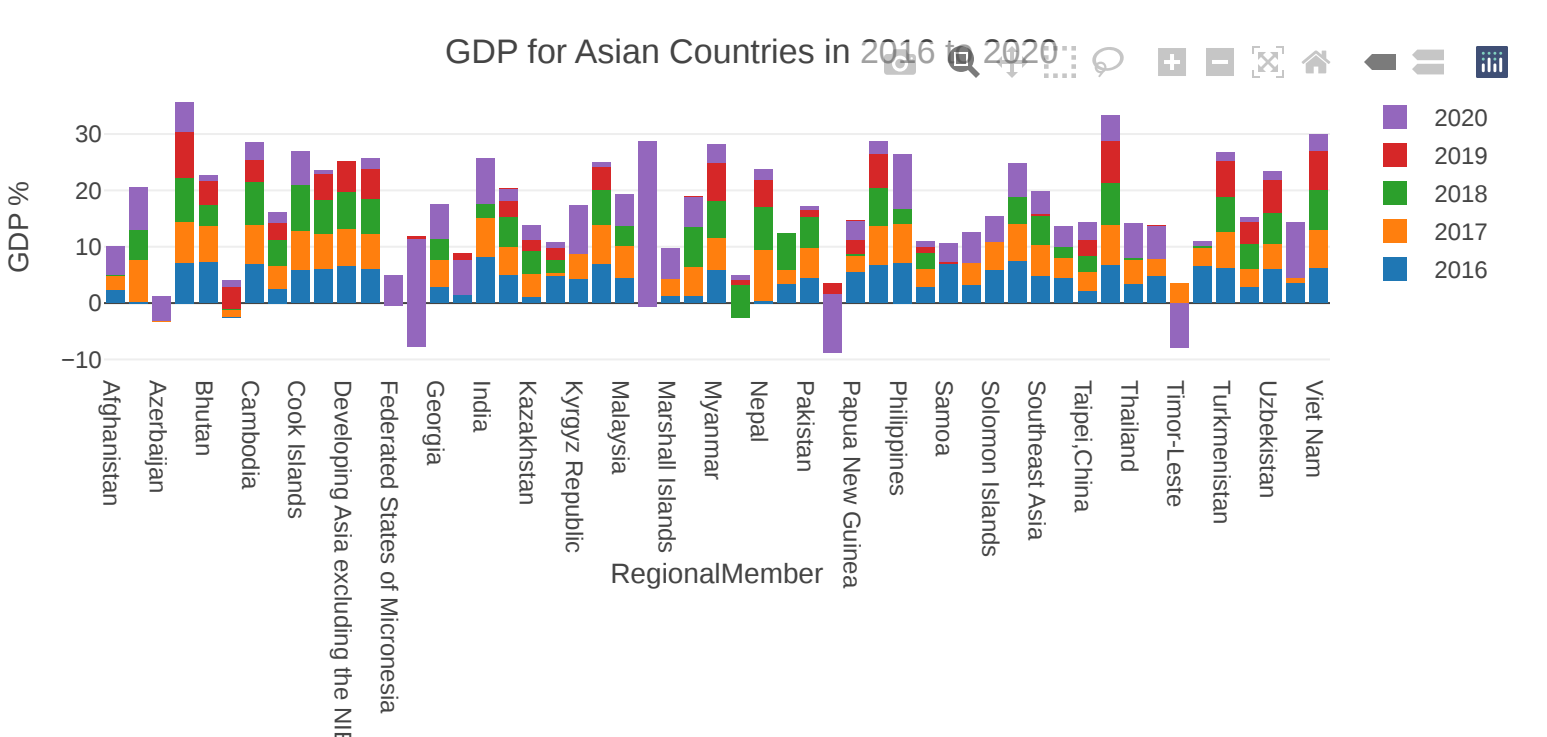
The pie chart below shows how many countries are present in each subregion of Asia. The Other category includes smaller parts of Asia that are still developing and may not necessarily pertain to any established country or subregion From this data we can see that a majority of Asian countries belong in either Southeast Asia or the Pacific, which equates to a total of 50.9% of Asia, while the still developing(other) countries in Asia only make up a small 3.77%, and lastly there is an even distribution of countries in both South and Central Asia, equalling 17% each.

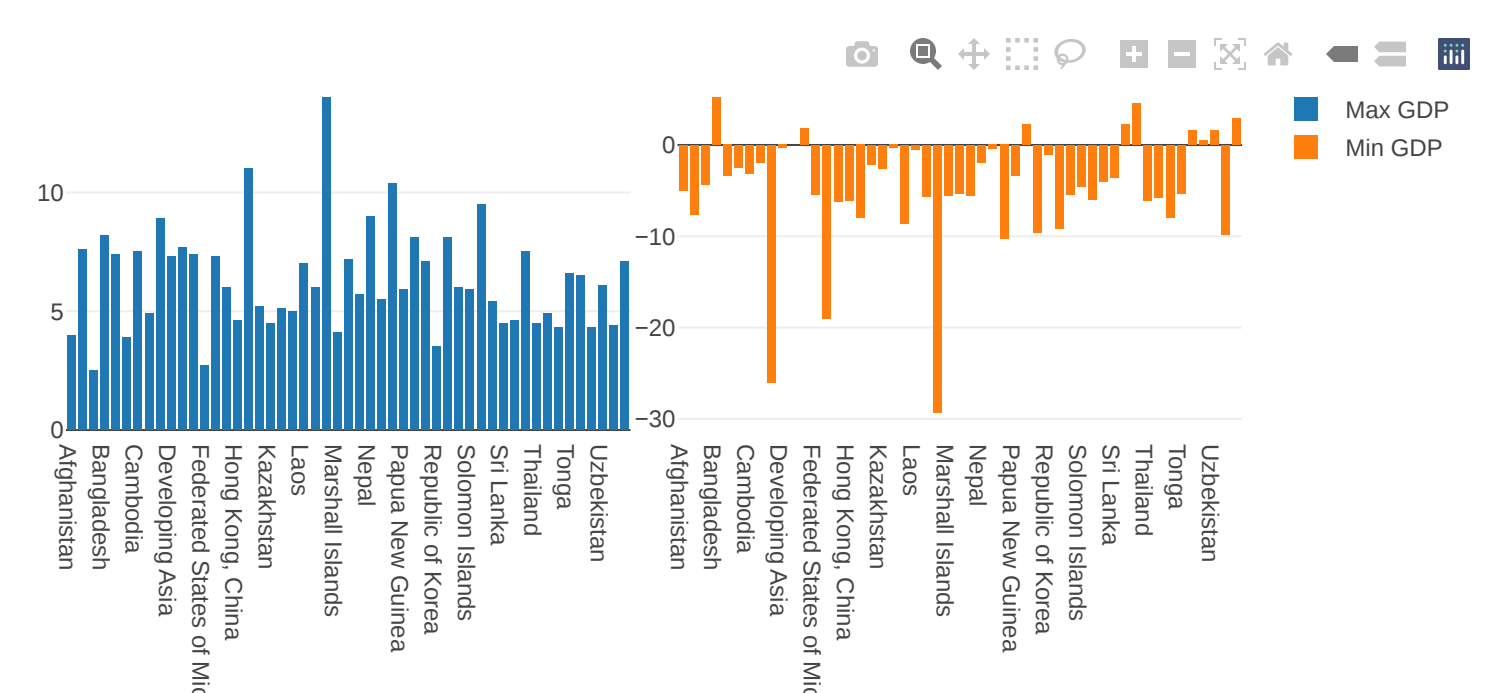


Change in GDP over a 5 year period

The below chart shows the GDP activity of differing Asian countries from the year 2016 to 2020. From this data it can be seen that the countries with very minimal economic activity are Azerbaijan, Brunei, Micronesia and Nauru. The countries with drastic changes to their GDP include Bangladesh, India, The Maldives, China, Philippines, Tajikistan, Turkmenistan Uzbekistan, and Vietnam. Also during this time period we can observe what the highest and lowest growths were for each of these countries, with the Maldives reaching 14% growth, but also reaching -29.3% growth.

It can be seen that during this time period a large majority of countries have experience some kind of negative growth. However there are also some countries that despite their neighbors not doing as well manage to maintain a positive growth rate throughout this entire period such as Bangladesh, China, Vietname, and Tajikstan.



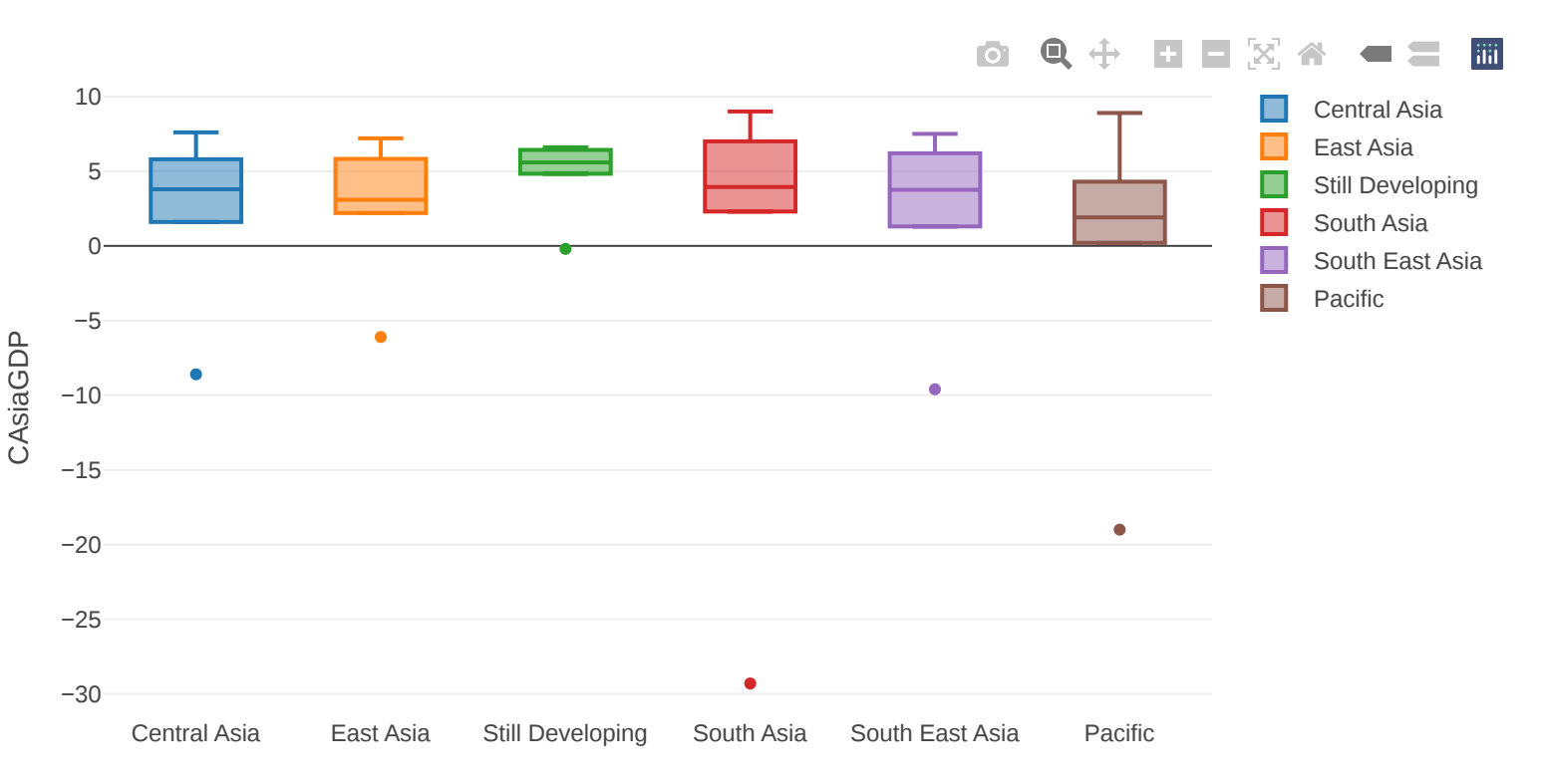


Box plot analysis of differing subregions of Asia

The below is a table of the five number summary for each subregion of Asia over the course of 5 years. From the table it can be seen that the still developing countries(Other) in Asia on average has the highest GDP growth. This makes sense because newly industrialized countries in theory should have the highest growth rate because they are still developing and have a lot of leeway.

Outside of still developing countries we notice that the average GDP growth for central, south, and south east Asia are all very similar, except for their outliers Lastly from the data we see that the pacific has the lowest GDP growth, which makes sense because it is difficult for a mass of islands to develop economically.

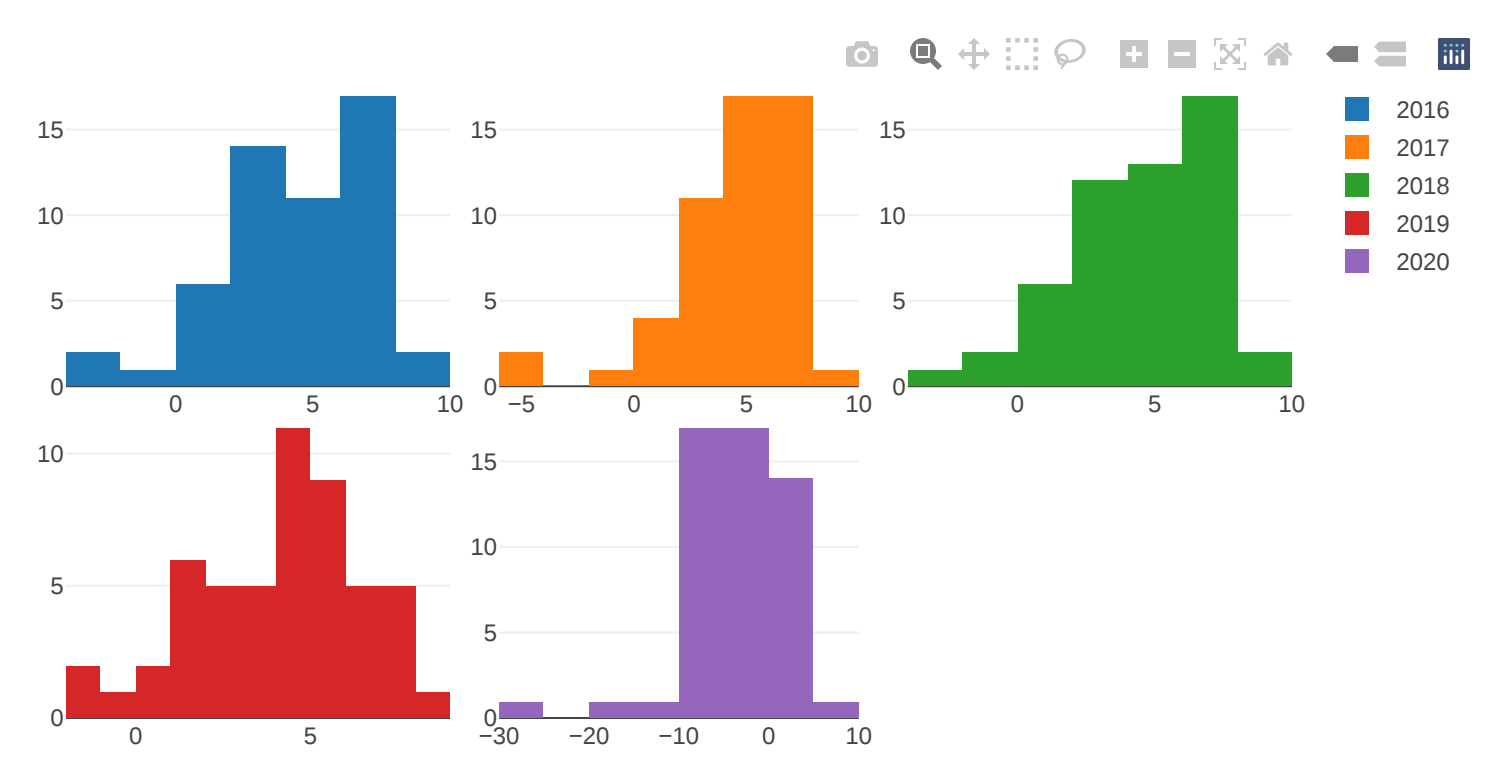
```
##      CentralAsia EastAsia Other SouthAsia SEAsia Pacific
## Min         -8.6     -6.1  -0.2     -29.3   -9.6   -19.0
## Q1           1.6      2.2   5.1       2.3    1.3     0.2
## Q2           4.5      3.0   6.0       4.6    4.4     2.4
## Mean         3.1      3.1   4.8       3.3    3.1     1.4
## Q3           5.8      5.8   6.4       7.0    6.2     4.3
## Max          7.6      7.2   6.6       9.0    7.5     8.9
```

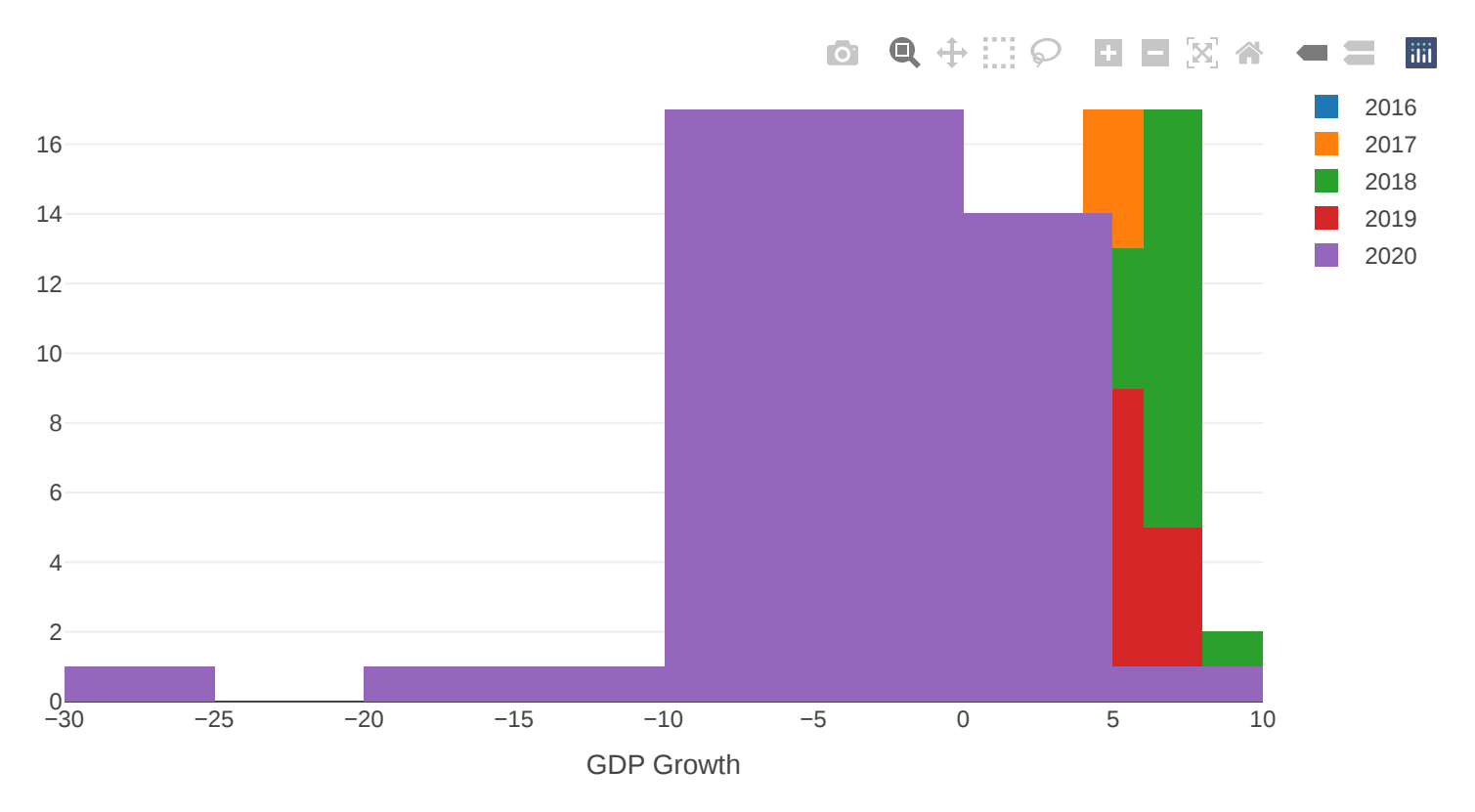


Analysis of GDP change over 5 years

From the different density of the histograms over this time period we can see how roughly on average how the growth rate of these countries in 2016 is around 5%-7%, and how from 2016 to 2018 the average growth rate either stays around the same or increases marginally. However on 2019 to 2020 we can see the average decrease from 5%-7% to 0-5% and then a further decrease to predominately negative values. This can be clearly seen on the aggregate histogram where the highest frequency peak shift from right to left.

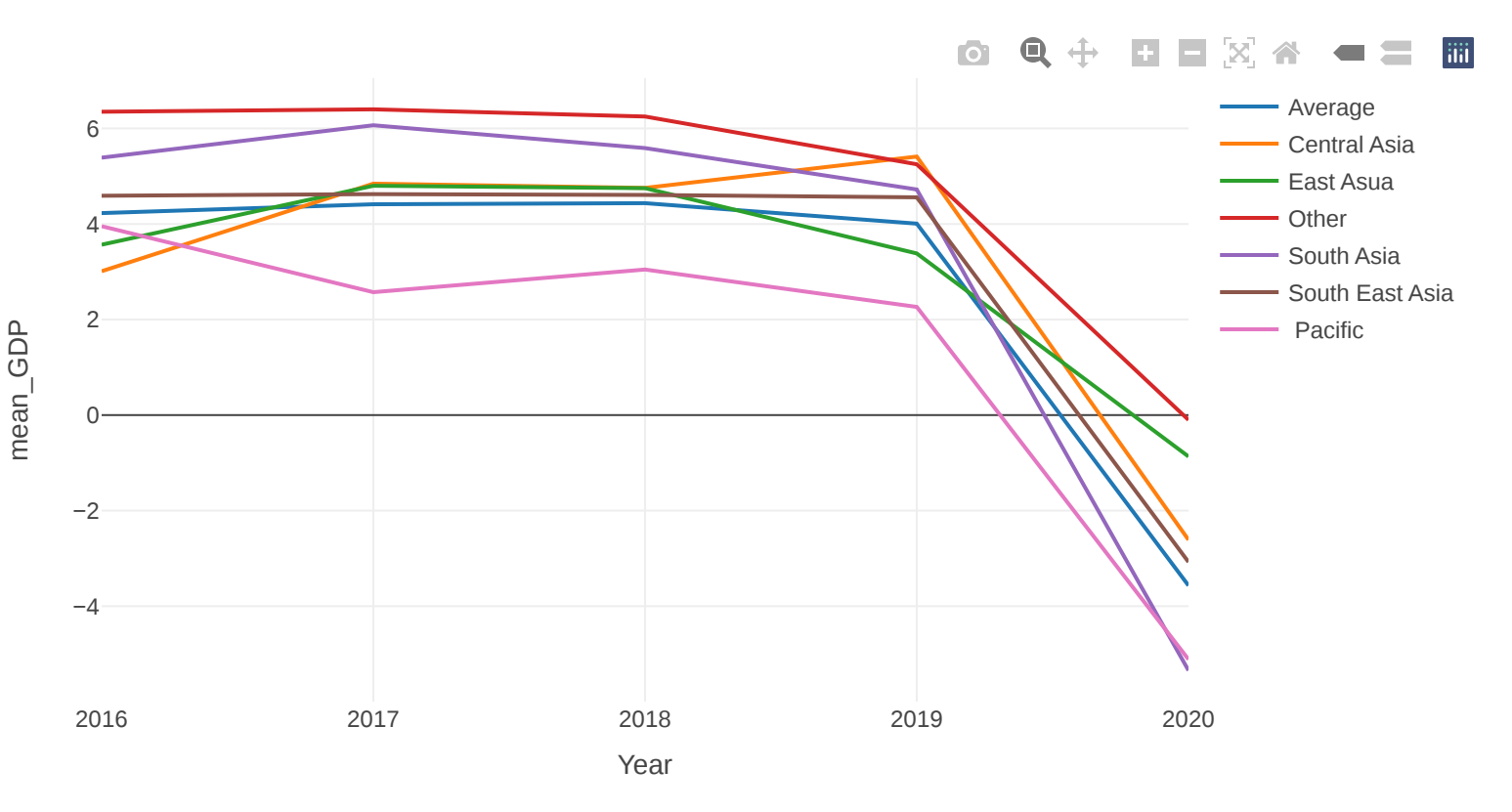
This is around the same time covid broke out and hit its peak, while this may not be the only factor determining GDP it is not a coincidence that there is a correlation between the outbreak of covid and the economic growth of these countries.





The previous histograms took a at the economic state of each individual country. This part examines the overall trend for each subregion in Asia, to allow for a greater overview into the overall trends going on in Asia. From the progression of the average GDP across the different subregions of Asia we can see that starting at 2018 the negative effects of covid on the economy become more apparent as time goes on, accumulating into a sharp GDP drop from 2019 to 2020.

For the years prior to 2018 we can see that each of the subregions have maintained a steady growth of around 4 to 6%. In this time frame it can be seen that the still developing areas(other) have the highest averages for growth across all 5 years, and the pacific has the lowest averages of growth across all 5 years. As for the other subregions of Asia their economic activity can be considered relatively similar



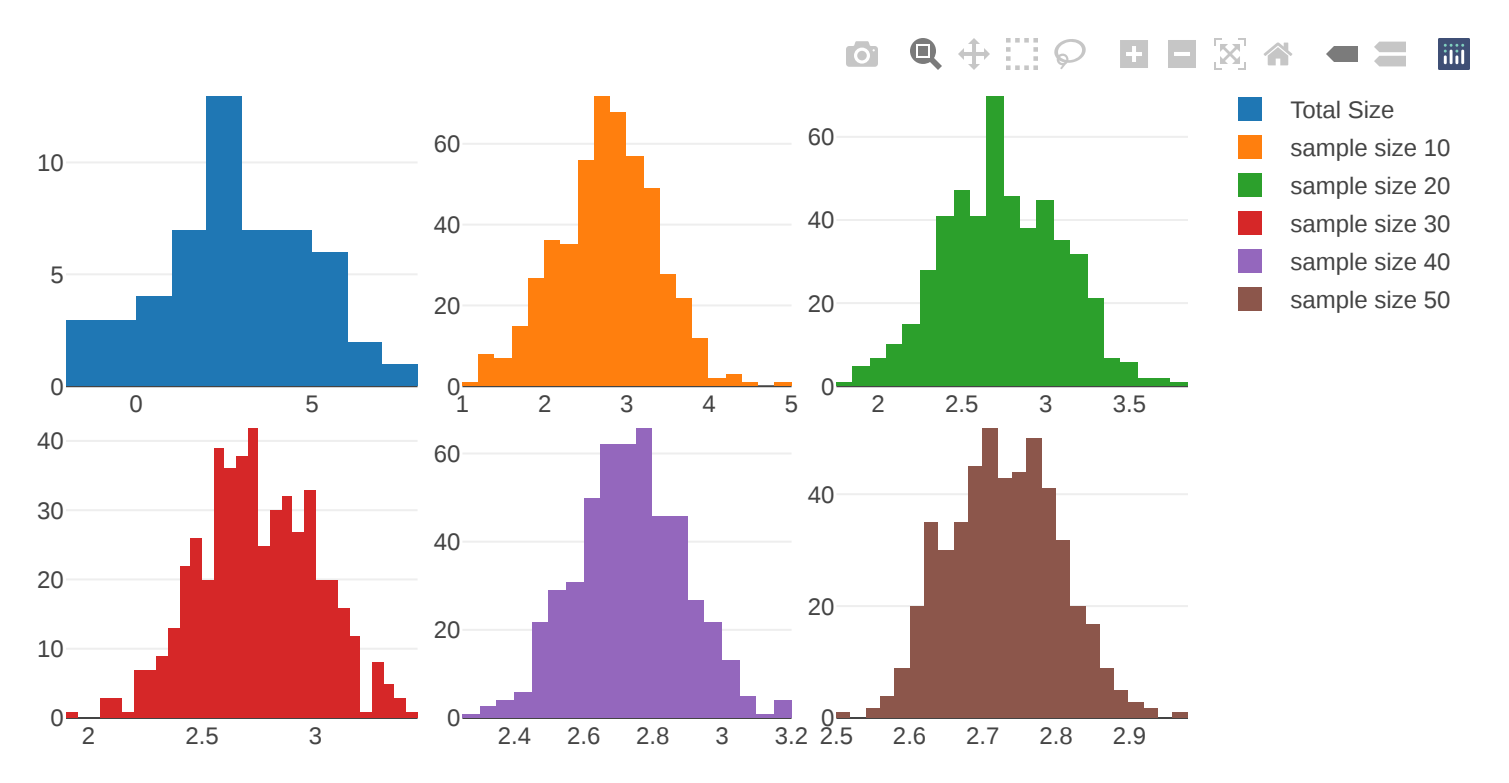
Applying the Central Limit Theorem

The central limit theorem dictates that the distribution of a sample means for any sample size has the shape of a normal distribution. Which means that as the sample size increases it should become closer and closer to that of a normal bell curve.

There are a few things we can see within this method. Firstly from the table below it can be seen that as the sample sizes increases, the standard deviation becomes closer and closer to 0 which means that the values become closer to the average. Because as sample size increases, the distribution approaches the normal distribution. As the sample size increases the mean should in theory also come closer to the normal mean, which can also be seen in this example, however due to the limited size of the total population variance is also a bit lower.

We can also see that as the sample sizes increase from the initial 10 to the final 50 the curve also becomes closer and closer to that of a normal curve. Which means that the average GDP of some countries will be lower and some higher, but a majority will be within the standard range.

```
##                StandardDeviation Mean
## Total                      2.098  2.7
## Sample Size 10             0.608  2.8
## Sample Size 20             0.358  2.7
## Sample Size 30             0.263  2.7
## Sample Size 40             0.159  2.7
## Sample Size 50             0.076  2.7
```

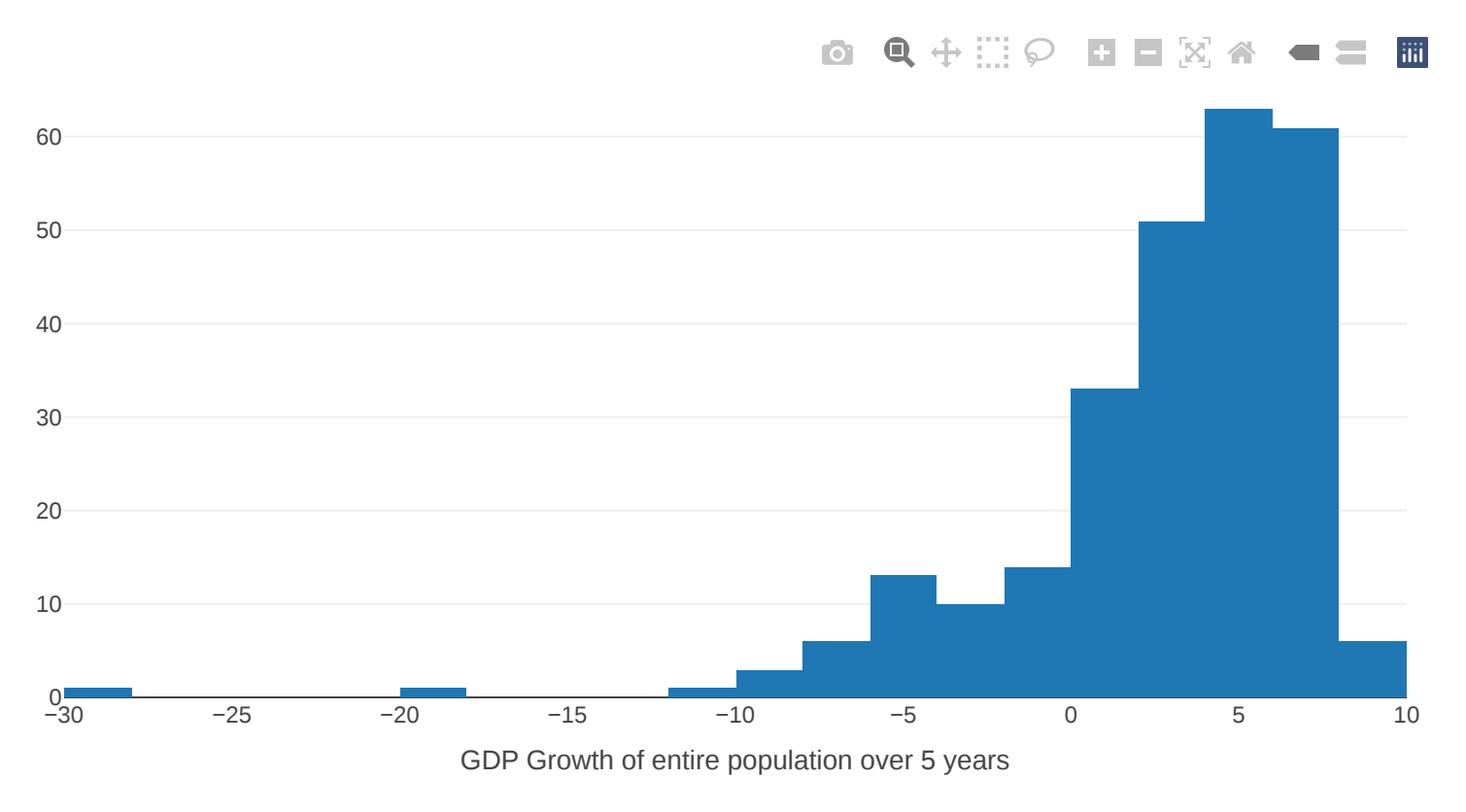


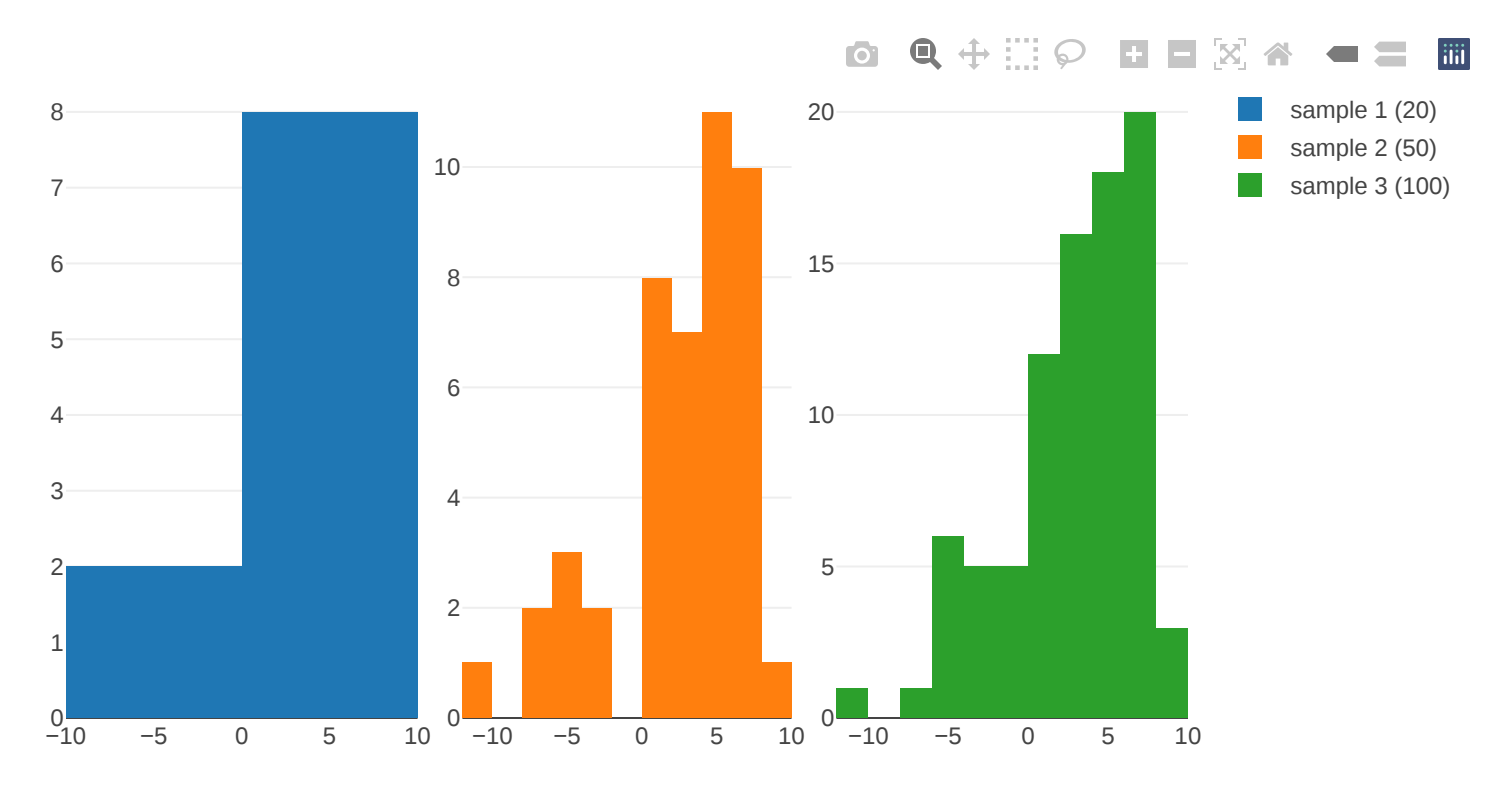
Sampling

Sampling is a method where a smaller portion of an entire population is selected and analyzed. The results form this analysis can be used to infer information on the overall population. Some different methods of sampling are simple random sampling, Systemic Sampling, and stratified sampling.

Random Sampling

The first method seen from the charts below is the simple random sampling method. From the sampling sizes below we can extrapolate a few things. As the sample size increases the mean becomes closer and closer to the mean of the entire population. This means that from a smaller sample size we can get a rough estimate of the entire population. This can be easily seen from the table below. Another piece of data that we can extrapolate is the overall trend of GDP growth. The first histogram is a plot of the entire data set, while the subsequent plots are of made up of varying sizes of samples. It can be seen that as the sample sizes increase it becomes closer and closer to the shape of the overall population. This means that samplings can be used to reflect overall populations.

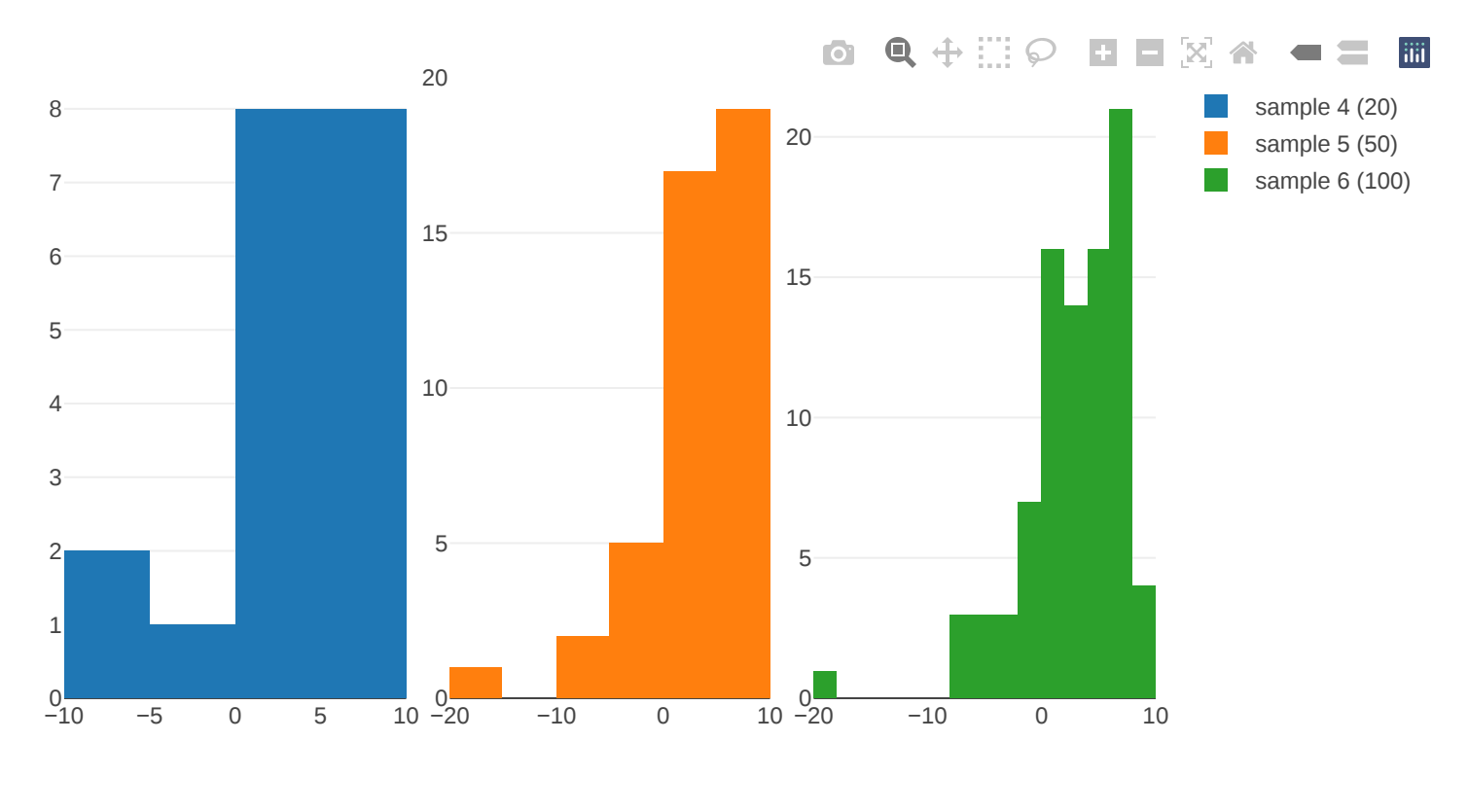





```
##            Mean StandardDeviation
## Population  2.7               4.7
## Sample 1    2.1               4.6
## Sample 2    2.5               4.6
## Sample 3    2.7               4.2
```

Systemic Sampling

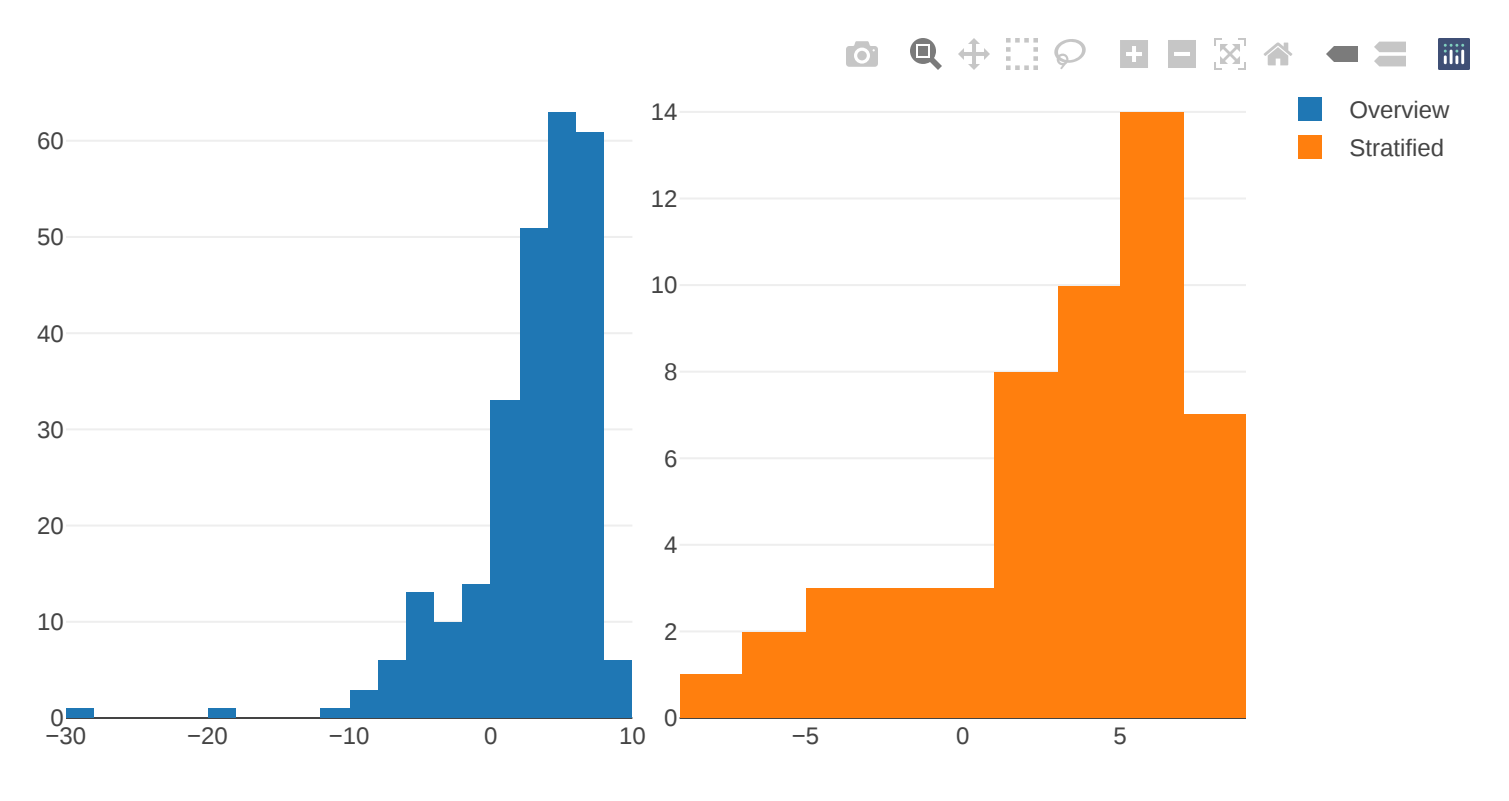
Systemic Sampling is a differnt of method sampling from simple random sampling. The way systemic sampling works is that it divides a population into different groups and a sample is taken systemically from each group. For example in sample 6 the entire population is divided into groups of 20, and from each sample taken from each group is in the same position as subsequent samples, meaning that if a sample drawn from the first group was position 3, the subsequent samples will also be from position 3 in all subsequent groups. When comparing the histograms from the Systemic Sampling to the population we can see that the results are similar to that of normal random sampling, and as sample sizes increase the plots become increasingly similar, which further proves that samples can be used as a representation of an overarching population.




```
##            Mean StandardDeviation
## Population  2.7               4.7
## Sample 4    2.1               4.6
## Sample 5    2.5               4.6
## Sample 6    2.7               4.2
```

Stratified Sampling

The last sampling method is stratified sampling, the population is divided into different stratas, and samples are drawn from each strata. From the plots we can also see how a stratified sampling can be indictive of entire populations.




```
##            Mean StandardDeviation
## Population  2.7               4.7
## Sample 7    3.1               4.0
```

Conclusion

The samples from the different sampling methods all show results that are indictive of the overall population. As sample sizes increase the standard deviation decreases and the samples becomoe progressively more indictive of the overall population. What this means is that a some discretion should be applied when determining the size of a usable sample size. Based on the different resultant plots both random sampling and systemic sampling are accurate enough to portay overall populations. Stratified may be less accurate but this can be attributed to either the entire population being too small, or limited diversity in the different stratas. Stratified Sampling can still be an accurate sample given enough information and a more diversified strata.