Project ReportIntroduction to Big Data - Term Project

Global Economic Monitor

April 19, 2018

Omkar Kakade  
Rohit Ravishankar  
Anika Sharma

# 

[**Introduction**](#_4lqp25cx7kth) **2**

[**Analysis**](#_xpeva8hd19hj) **5**

[**Conclusion**](#_i85ws8se9wc5) **13**

[**References**](#_tlenj1ut0e29) **14**

# 

# **Introduction**

# The Data

The dataset we have chosen is titled *“Global Economic Monitor”*.

The Global Economic Monitor is designed to provide readers with a comprehensive overview of the key macroeconomic views and forecasts. Its aim is to provide a global view with a particular focus on the interactions between mature and emerging economies. [1]

The dataset originally consisted of 41 files with data on 100+ countries. It is key to note that not all countries had data on all these indicators, hence the final choice of files was driven by this fact.

Some of the key indicators are listed below:

1. Core CPI,
2. CPI Price, % y-o-y, median weighted
3. CPI Price
4. Country
5. CountryNotes
6. Exchange rate, new LCU per USD extended backward,
7. Exchange rate, old LCU per USD extended forward,
8. Exports Merchandise, Customs, current US$, millions,
9. Exports Merchandise, Customs, current US$, millions,
10. Exports Merchandise, Customs, Price, US$,
11. Exports Merchandise, Customs, Price, US$,
12. Footnotes
13. GDP,constant 2010 LCU,millions,,
14. GDP,constant 2010 US$,millions,
15. GDP,current LCU,millions,
16. GDP,current US$,millions,
17. Imports Merchandise, Customs, constant US$, millions,
18. Imports Merchandise, Customs, constant US$, millions,
19. Imports Merchandise, Customs, current US$, millions,
20. Imports Merchandise, Customs, current US$, millions,
21. Imports Merchandise, Customs, Price, US$,
22. Imports Merchandise, Customs, Price, US$,
23. Indicators
24. Industrial Production, constant US$,
25. Industrial Production, constant US$
26. Nominal Effective Exchange Rate
27. Official exchange rate, LCU per USD,
28. Real Effective Exchange Rate
29. Retail Sales Volume,Index
30. Series
31. SeriesNotes
32. Stock Markets, LCU
33. Stock Markets, US$
34. Terms of Trade
35. Total Reserves
36. Unemployment rate,Percent

**Our final dataset was narrowed down to the below 5 key indicators as chosen from the list above :**

* GDP
* Exports Merchandise
* Imports Merchandise
* Stock Markets
* Unemployment Rate

# Data Collection:

The data was collected from World Bank Open Data website.

The link for the dataset is included as: <https://datacatalog.worldbank.org/dataset/global-economic-monitor>

Countries in focus

The countries in focus were chosen to classify within three groups. These three groups being in regards to each respective economy as Developed, Emerging, and Under Developed. The list of countries along with their given group to be classified in are given below:

* United States *[Developed]*
* United Kingdom *[Developed]*
* Canada *[Developed]*
* Finland  *[Developed]*
* China  *[Emerging]*
* Brazil *[Emerging]*
* Estonia *[Under Developed]*
* Mexico  *[Under Developed]*

Data Preparation and Connection between SSMS and R

The following steps were observed and performed for data preparation :

1. Data on the CSV files were ***cleaned***to replace NA values as NULL values.
2. ***Import*** the data into SSMS and create corresponding tables, initially importing the data from CSV.
3. ***Install***ODBC driver and create DSN (Data Source Name) for connectivity to R.
4. ***Connect*** SSMS to R using the DSN and the project database file.
5. ***Transposing*** of the dataset such that the countries are represented as rows and years as columns. Initially, the data set showed countries as columns.
6. Column names were ***missing*** while importing data into SSMS. This was subsequently fixed by modifying under the design tab of the tables.
7. ***Imputed*** the missing values in the tables, with the mean of the data. This was done using the Data Cleaning transformation technique available in Rattle.

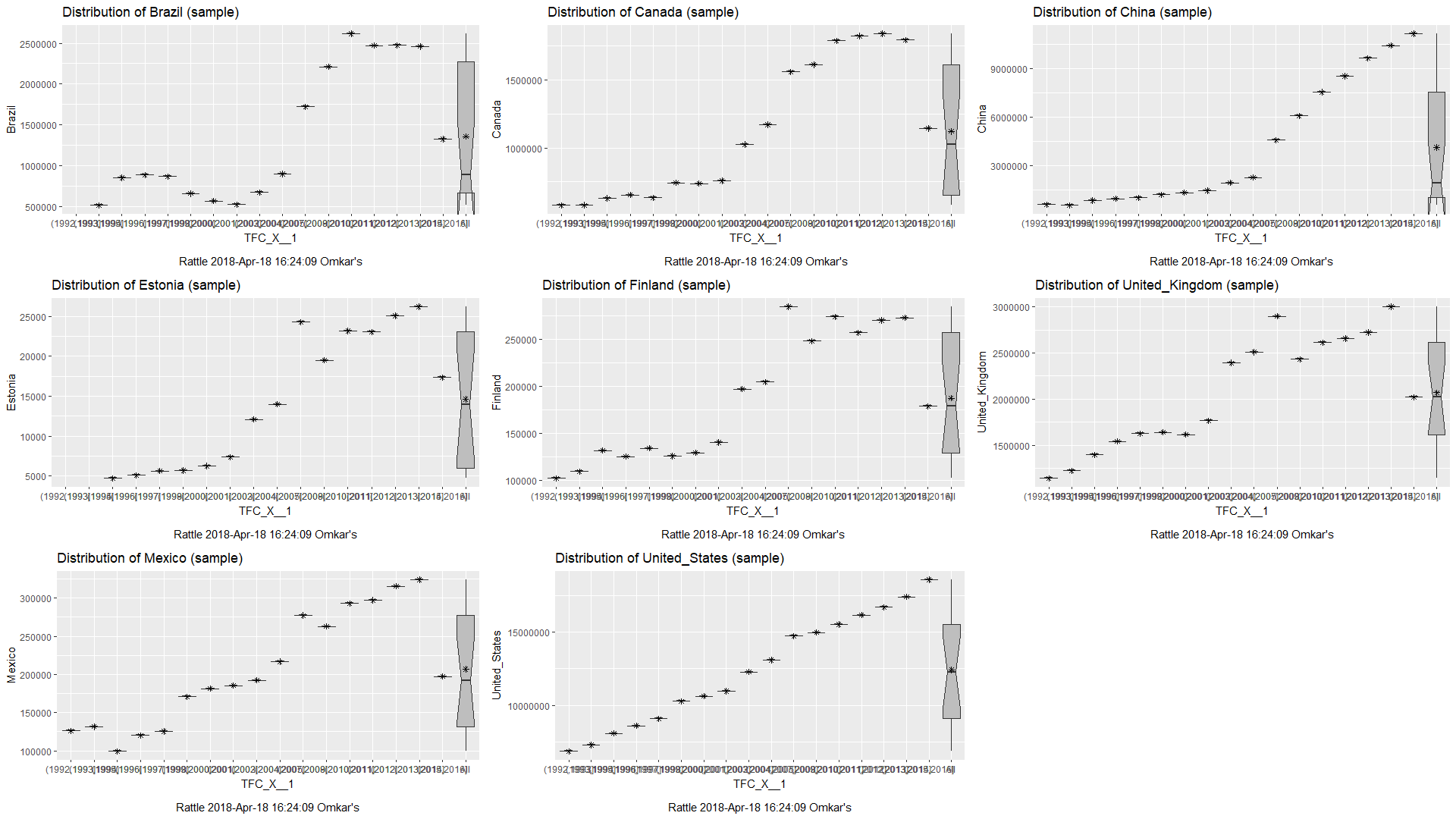
# 

# 

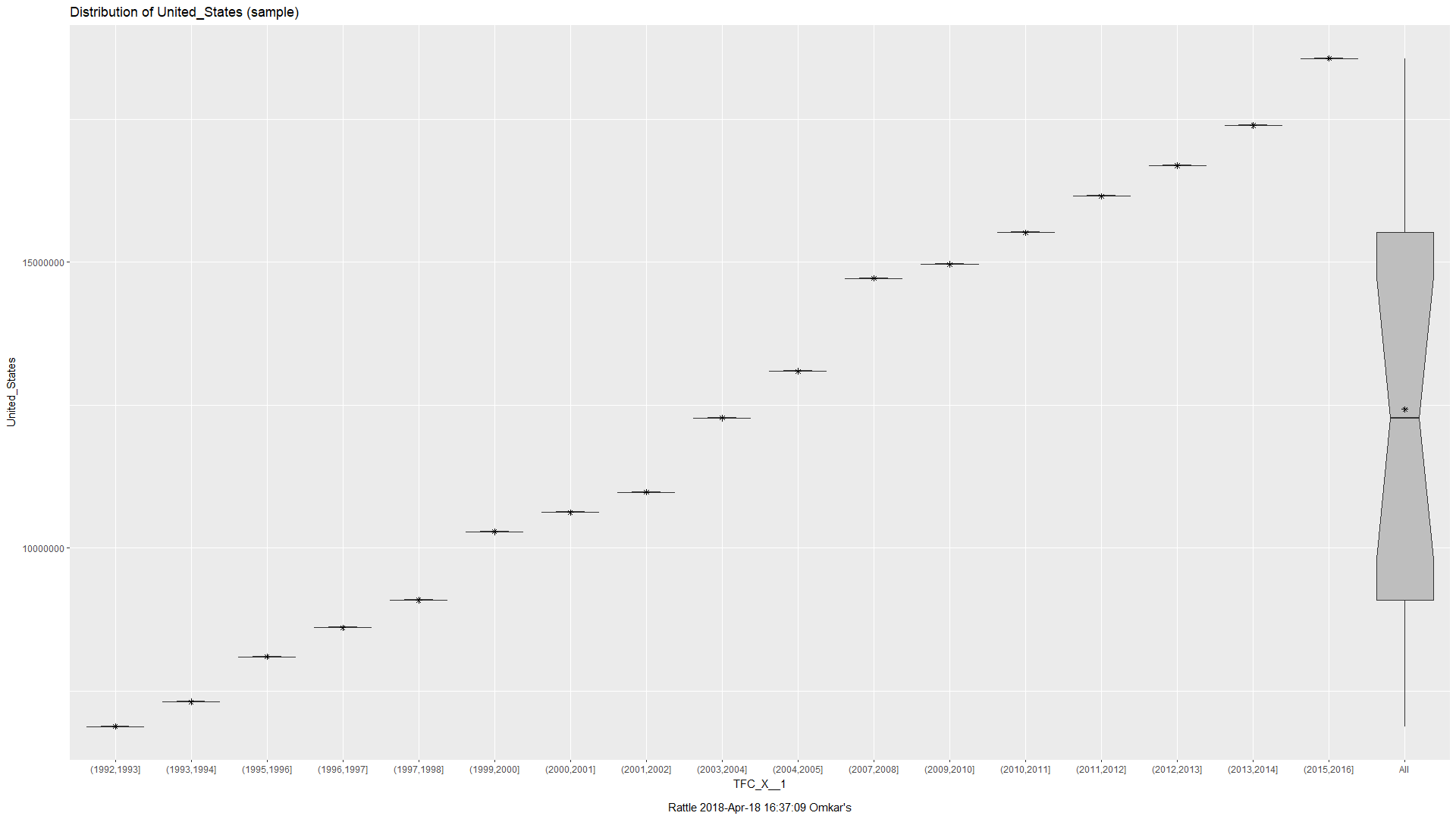
# **Analysis**

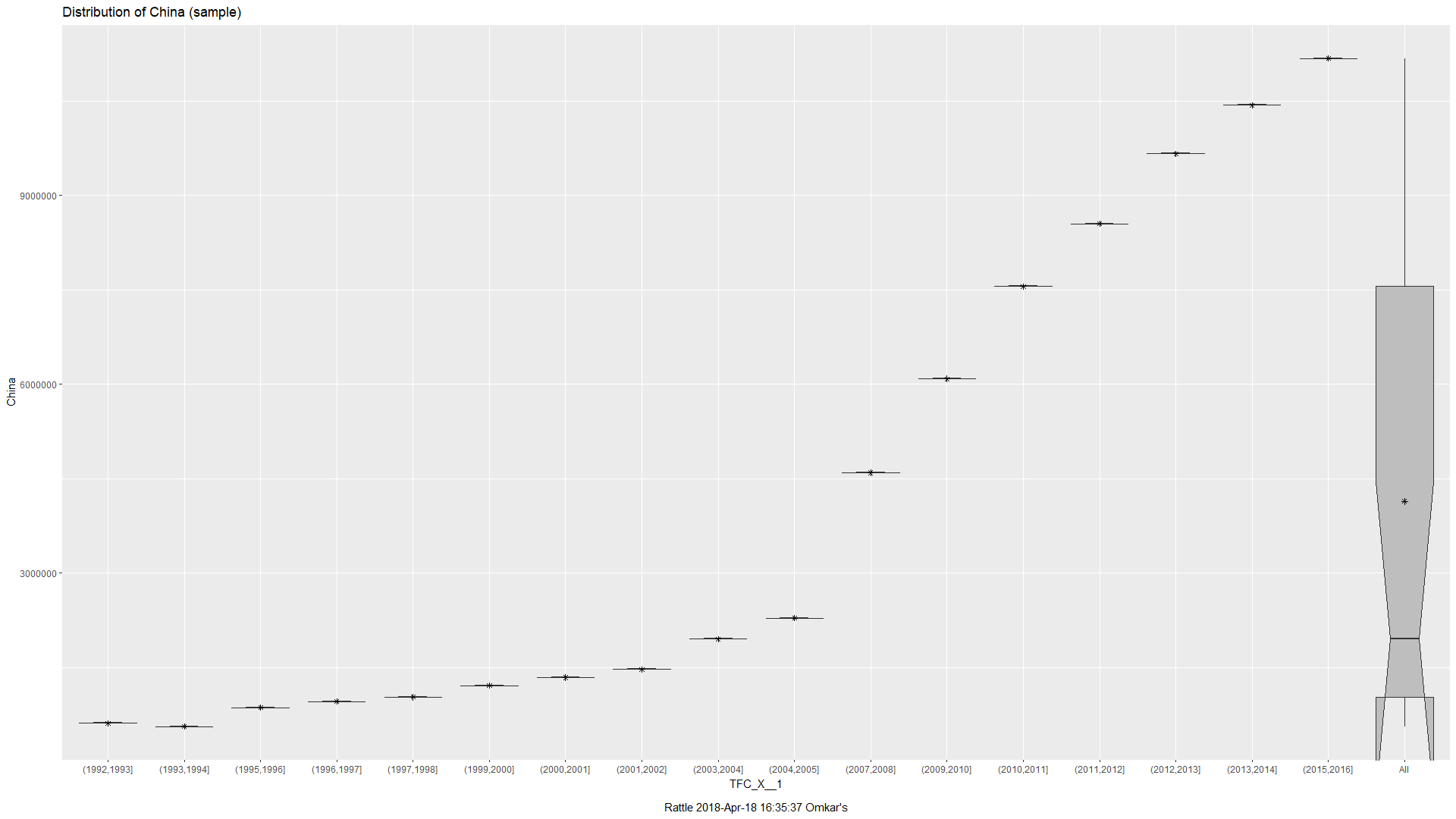
GDP - Gross Domestic Products

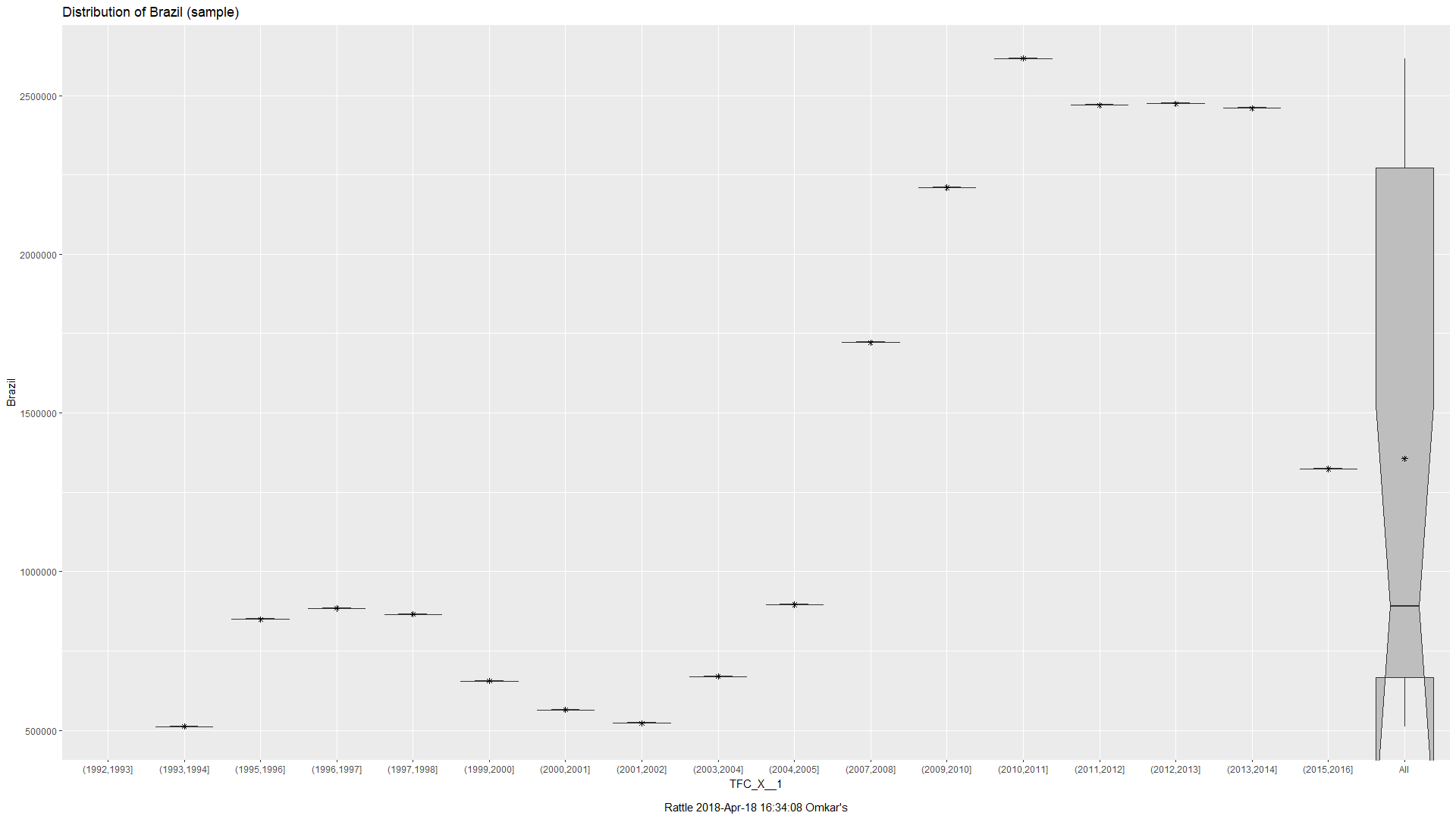
GDP is the first indicator we want to take a look at.   
Gross domestic product (GDP) is a monetary measure of the market value of all final goods and services produced in a period (quarterly or yearly) of time. Nominal GDP estimates are commonly used to determine the economic performance of a whole country or region, and to make international comparisons. Nominal GDP per capita does not, however, reflect differences in the cost of living and the inflation rates of the countries. [3]

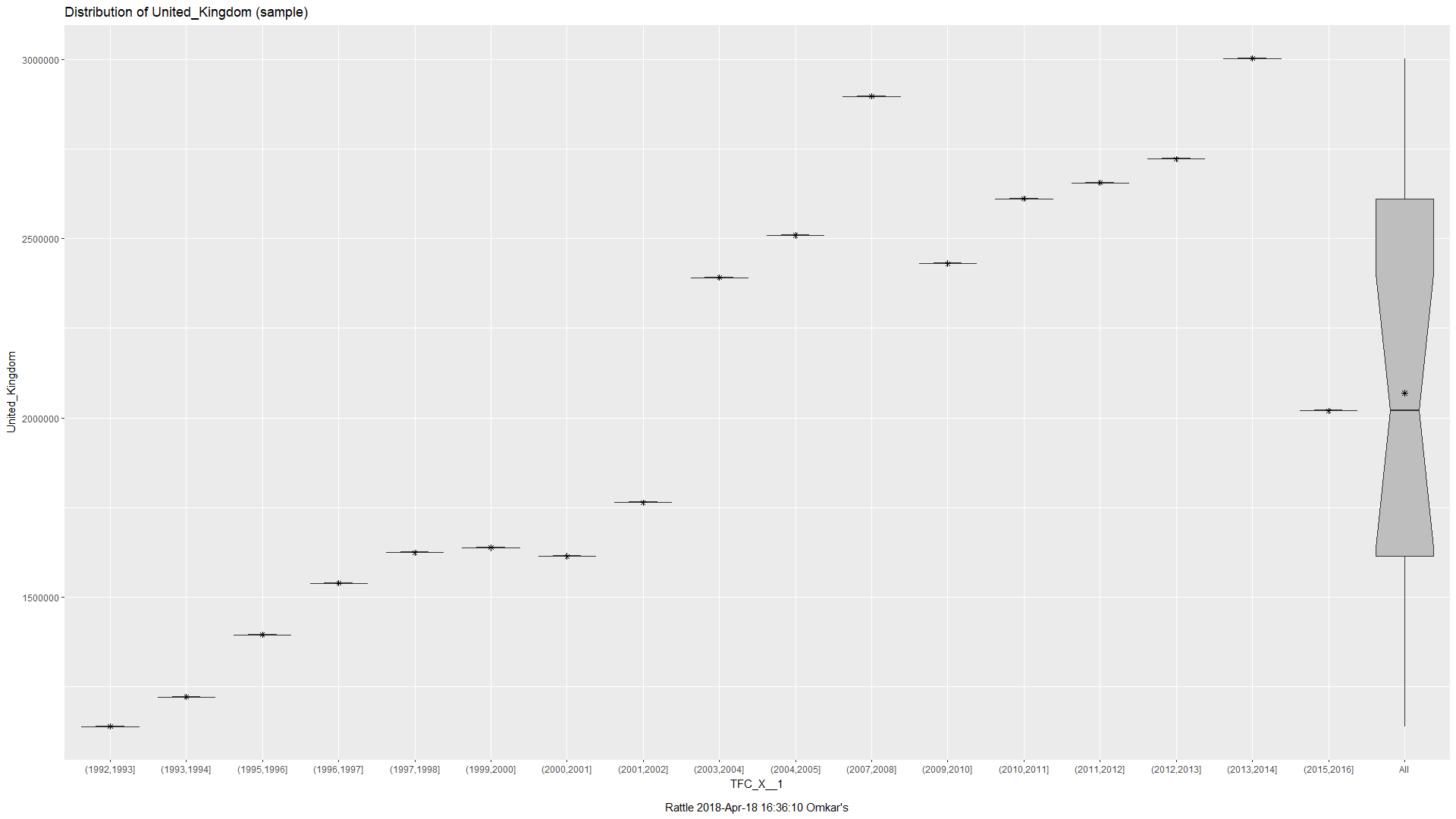
Analysis is done on the data of the above mentioned countries and a trend is established year on year from 1990-2017. (Please zoom in for all the plots)  
  


The above charts show the GDP year on year for 8 countries and the following findings are noted below -

The United States has enjoyed a steady and almost predictive growth over the 27 year period. There were significant events in 2008 and 2002. While these slowed down growth, they did cause any impede the economy.  
  
  
  
  
  
  
  
  


China experienced a burst of growth from 2007 onwards.   
  
  
  
  
  
  
  
  
  


Brazil experienced exponential growth starting from 2006, and was constant post 2011.  
  
  
  
  
  
  
  
  


UK experienced steady growth and a spurt post 2003 but a dip post 2008, which could be accredited to the 2008 Financial Crisis.  
  
  
  
  
  
  
  
From the above plots, China stood out the most specially for its spectacular growth post 2007, we suspected a correlation between other factors that might have led to this growth and decided to look at the Exports and Imports data for China for the same period.

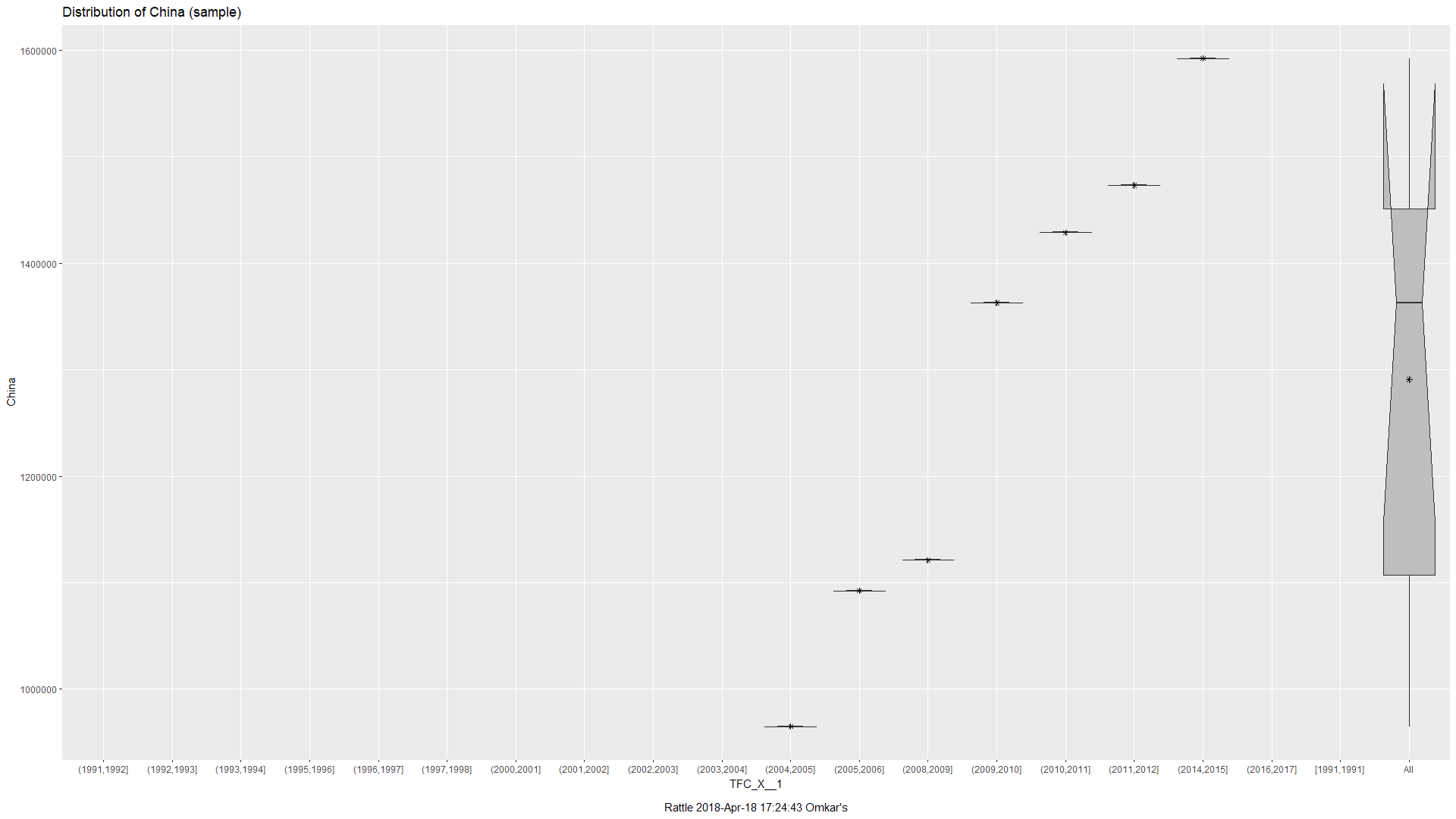
China Imports vs. Exports

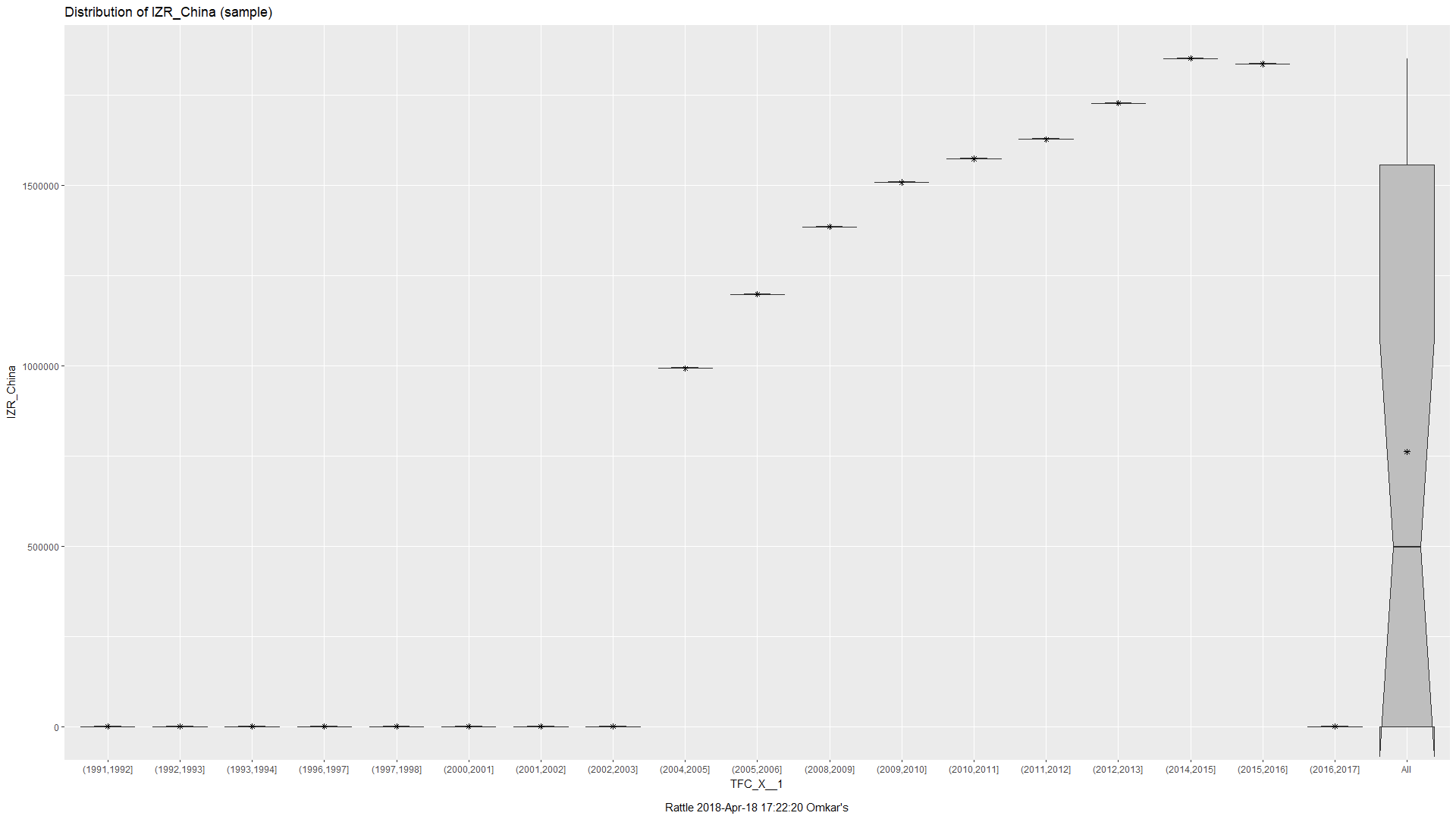
Export merchandising is a method of offering retail goods for sale in a foreign consumer market. Exports merchandise (measured in USD) is the total sum of the value of goods exported out of a particular country.[4]

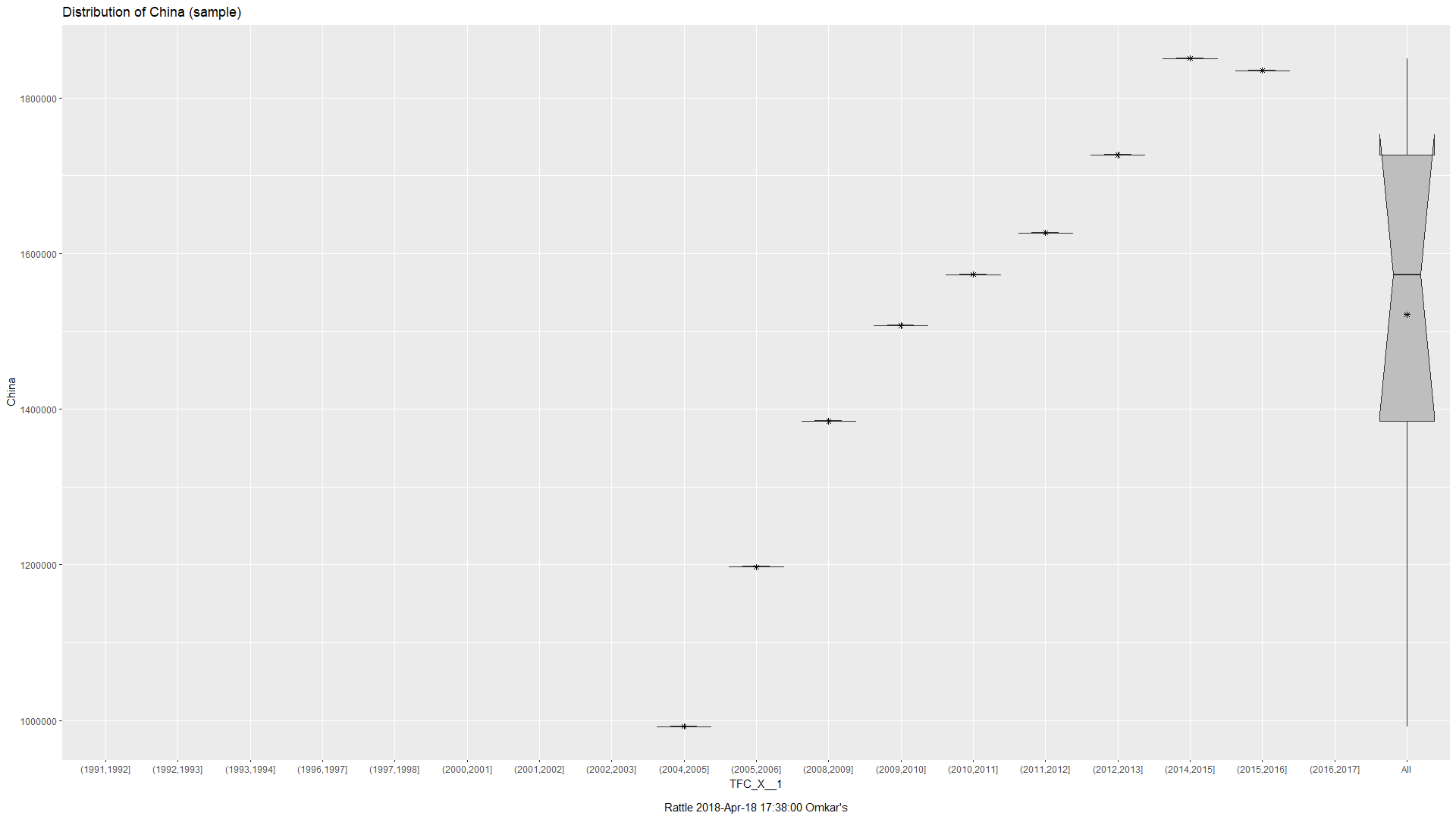
Import merchandising is the method of importing retail goods for sale in the country. Import merchandise (measured in USD) is the total sum of the value of goods imported into a particular country.

The following images depict the following, the first image depicts the Import Merchandise value of China in USD and the second image depicts the recoded version of the same. The third image depicts the Export Merchandise value of China in USD.

From the images below, it is clearly visible that China exhibits an export to import surplus. This has been an important factor that has helped the Chinese GDP grow year-on-year, at a rate which very few countries have been able to achieve.



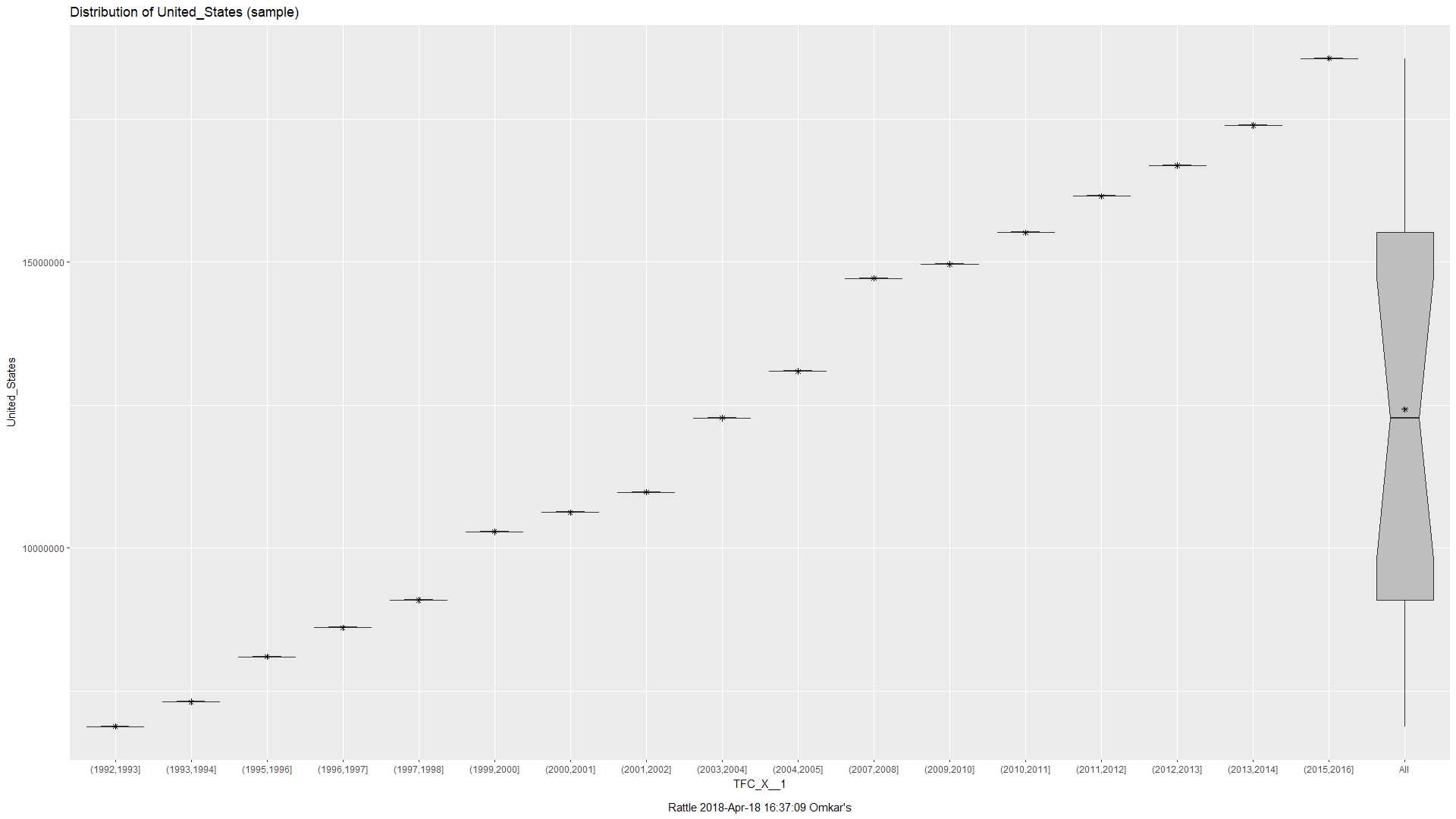


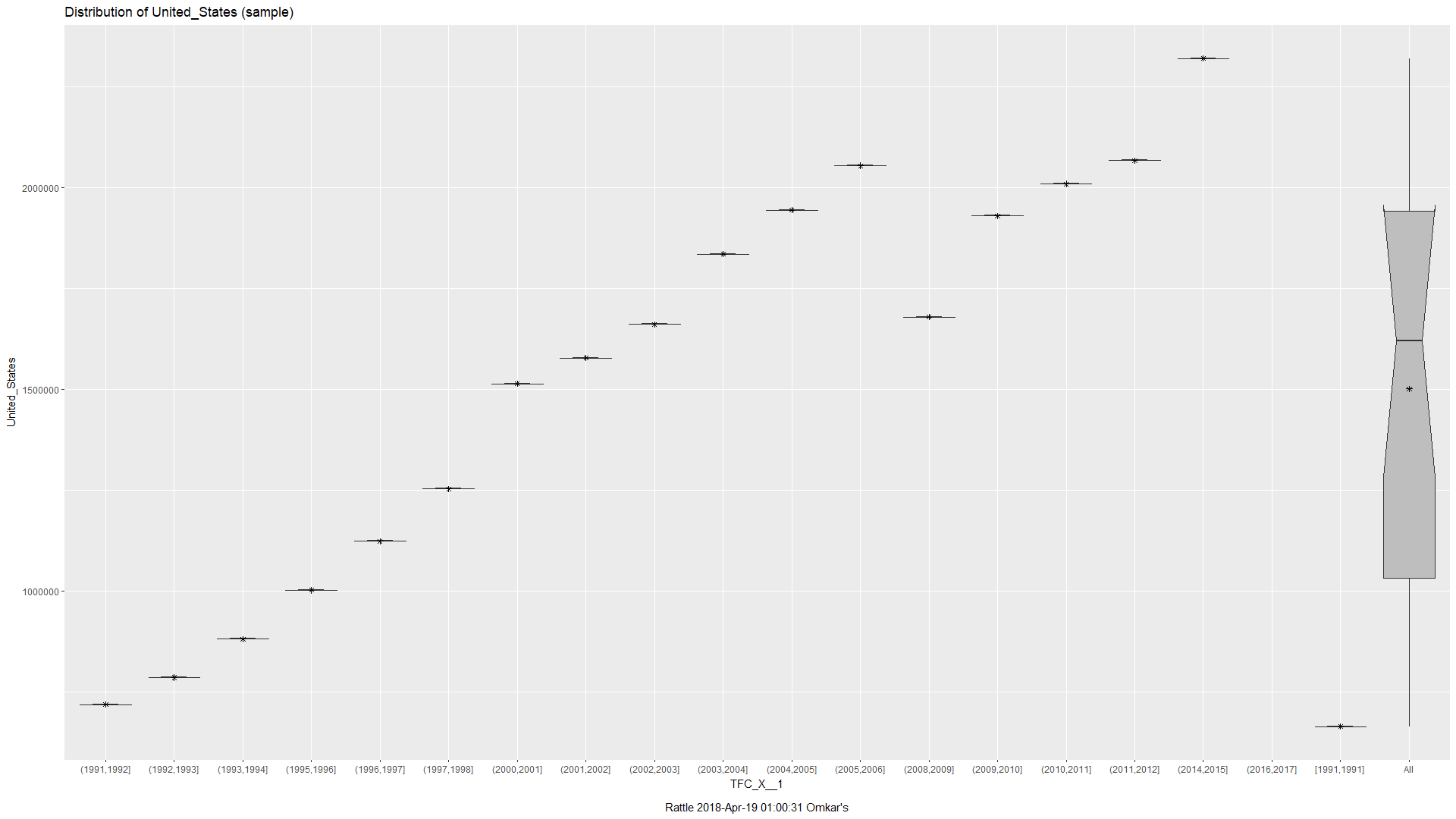


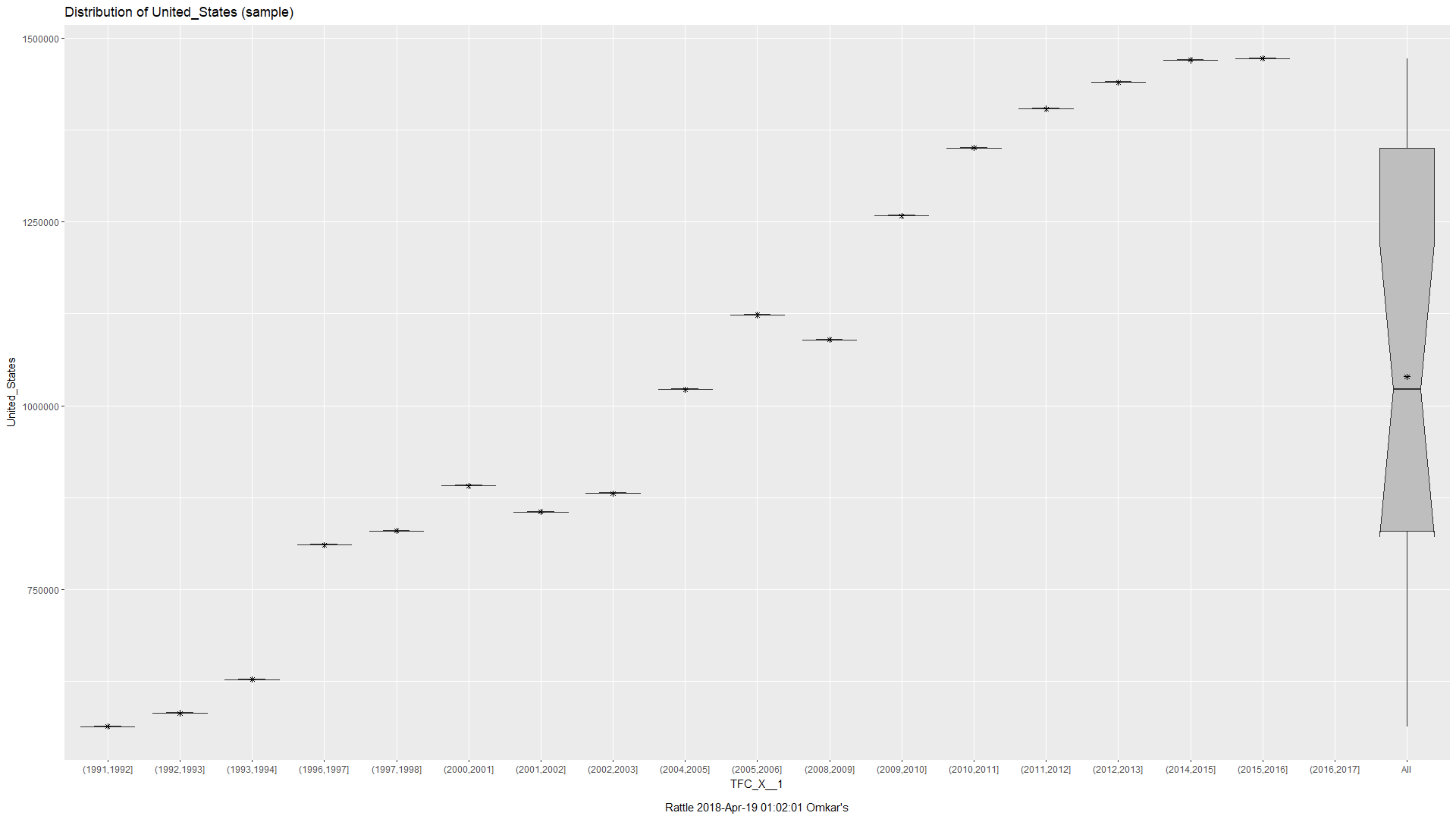
USA GDP vs Import

During financial meltdown of 2008, we notice that many of the economies like that of UK, Estonia and Finland were negatively impacted. US on the other hand, though impacted managed to grow at a steady rate.

A probable explanation for this could be that US decreased its dependence on imported goods and managed to keep the export at a steady rate thereby mitigating the impact.



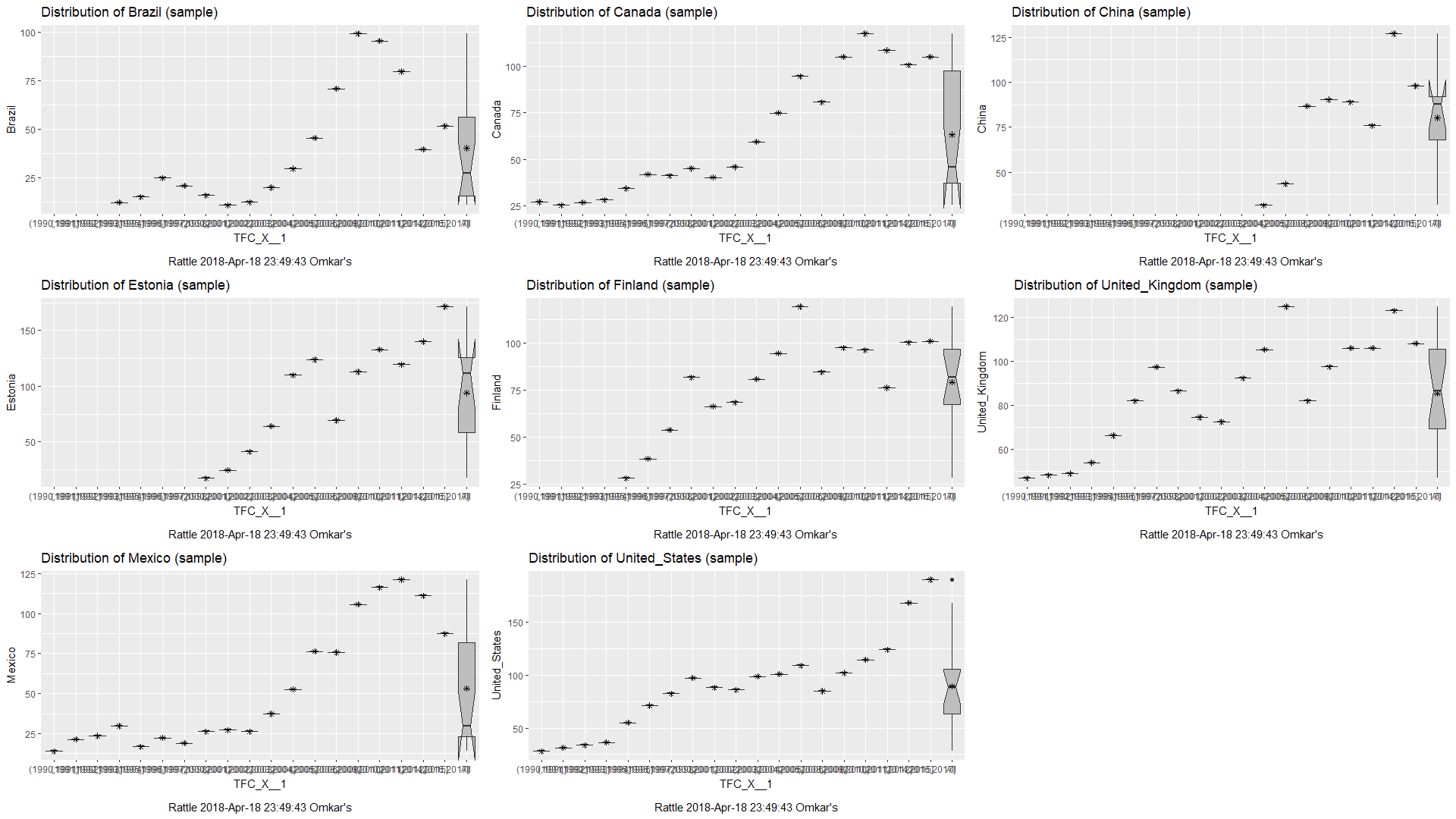


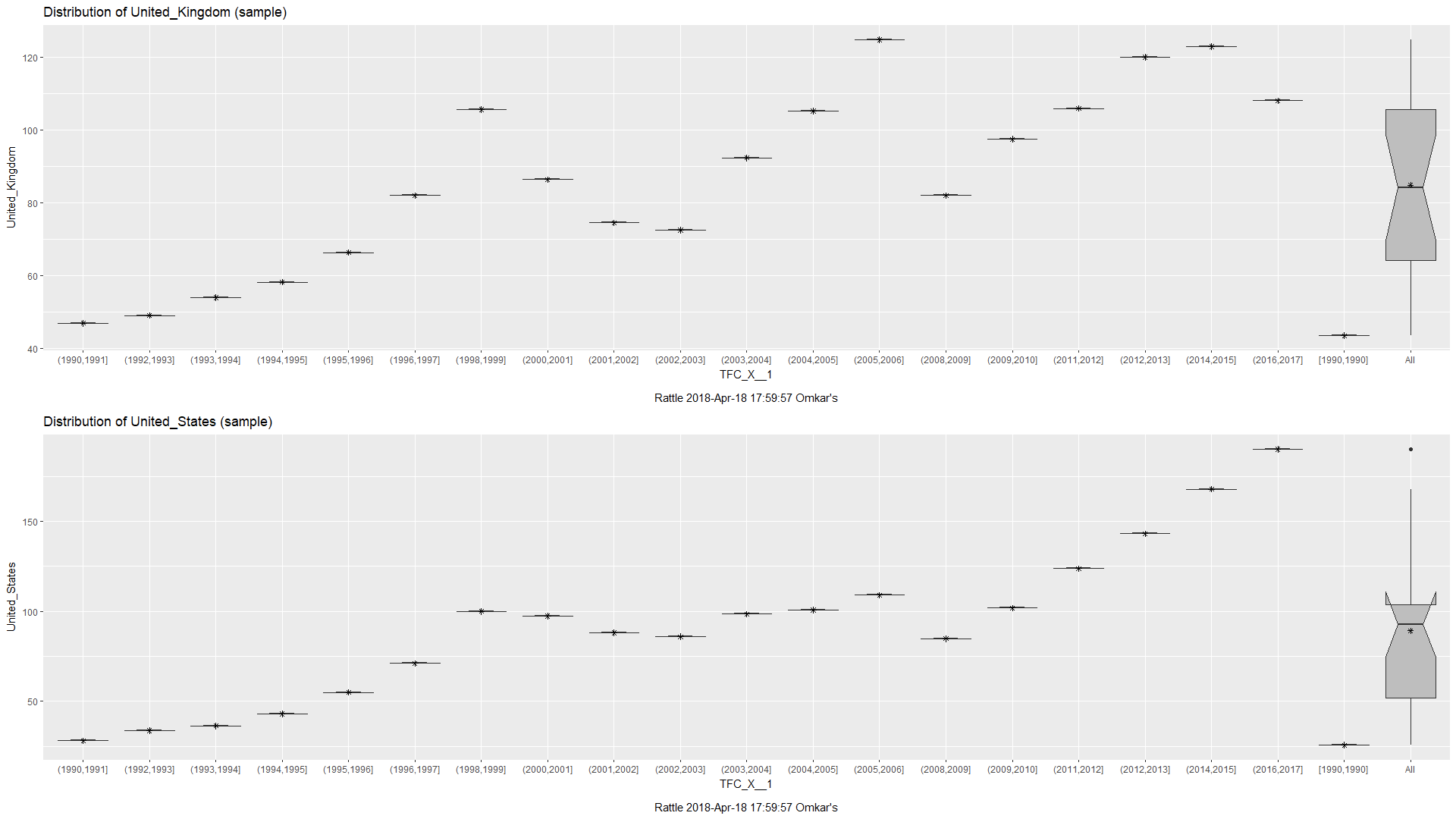


# Stock Market

An interesting trend we see is that there is a decline in the stock market valuation or decline in the growth of the stock market during the period of 2007-09. This was sparked by the financial crisis of 2008. It began with a crisis in the subprime mortgage market in the United States, and developed into a full-blown international banking crisis with the collapse of the investment bank Lehman Brothers.

Another interesting trend we spot with the stock market data is the decline in the valuation around the period of 1999-2002 in the developed countries such as US, UK and Canada. This was fueled by the dot-com bubble burst. It was a historic economic bubble and period of excessive speculation and a period of extreme growth in the usage and adaptation of the Internet.



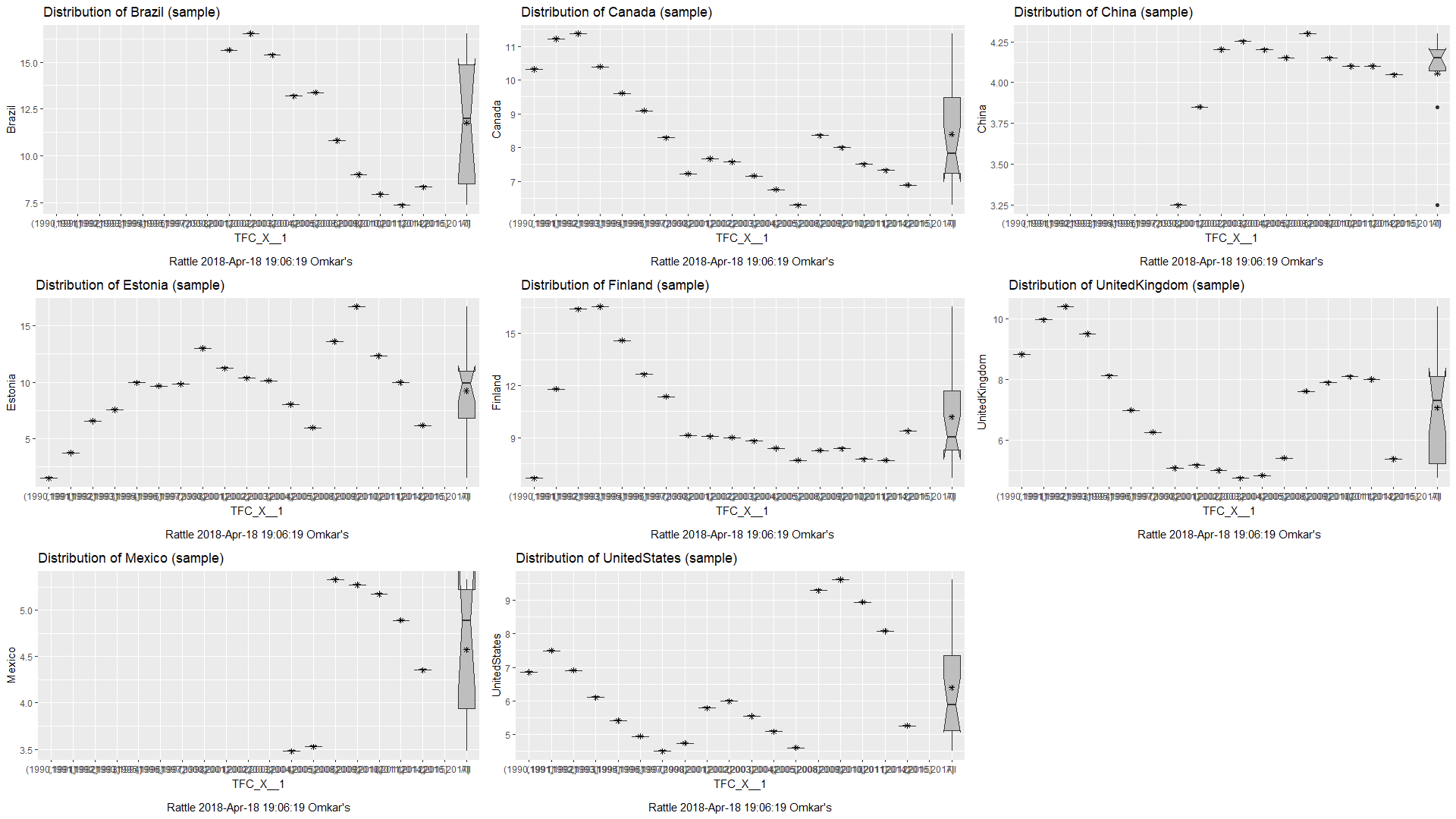




# Unemployment Rate

Akin to the stock market trend, we see is that there was steep rise in the percentage of unemployed people almost across all countries during the financial crisis of 2008. With investment banking firms such as Lehman Brother filing for bankruptcy, there was a huge loss of jobs and a cascading effect across the world.

Similarly, with the dot-com bubble crash as well, we see a steady loss of jobs leading to an increase in the unemployment rate in US, UK and Canada.



Another interesting insight to note is we see a steep decrease in unemployment rate in Brazil. This can be attributed to the increase in exported good from Brazil which helped increase employment across Brazil.

# 

# 

# **Conclusion**

In conclusion, we are able to obtain a data set and perform multiple operations upon the data in order to achieve some type of understanding of what the data represents. With our data set representing that of multiple economies across the globe, we accomplished multiple observations. These observations pertaining to those of the GDP, Imports vs. Exports, and Stock Markets of each respective countries’ economy during a time period of financial crisis and downfalls.

In order to achieve such examinations as we have performed through the analysis of this report, we must also acknowledge the steps we took in order to arrive at this stage. Our analysis was wholly contingent upon proper Big Data concepts and practices. These data operations including that of data cleaning, preparation, and database management. With these data functions available through that of R, Rattle, and SSMS as our main tools of use, we were able to display the goal of this project.

The goal of this project being to perform analysis such that we attain something meaningful from the data set. As we have observed from this report, we have concluded many different main points from the global economy data set.

# **Individual Contributions**

* Data Collection was handled by Anika Sharma.  
  The activities performed were
  + Searching for a reliable data source.
  + Finalizing the specific dataset.
  + Cleaning the data.
* Data Preparation was handled by Rohit Ravishankar.
  + Feed the cleaned data into SSMS.
  + Connect SSMS with R using ODBC.
  + Perform SQL Queries on the data.
* Data Analysis was handled by Omkar Kakade.
  + Transform the data in Rattle.
  + Deriving insights from the data with R.
  + Visualizing the data.

# 

# **References**

[1] Institute of International Finance, <https://www.iif.com/publications/global-economic-monitor>

[2]Kaggle- World Development Indicators <https://www.kaggle.com/worldbank/world-development-indicators>

[3] Gross Domestic Product wiki, <https://en.wikipedia.org/wiki/Gross_domestic_product>

[4] Export Merchandise <http://smallbusiness.chron.com/export-merchandising-16068.html>