

Title: Impact of COVID-19 on food production and Humans

Project report

Students:

Santosh Adabala, siad4700@colorado.edu

Evan McCormick, evmcc5552@colorado.edu

Anila Reddy Musku, anmu4597@colorado.edu

Yamini Priyanka Kandikattu, Yamini.kandikattu@colorado.edu

Maxwell Solko, maso1345@colorado.edu

Sushil Deore, sude8287@colorado.edu

Roja Kuchipudi, roku6452@colorado.edu

Vamsi Gontu, vago4483@colorado.edu

Sravani Rao Savaram, srsa1680@colorado.edu

Professor:

Dr. Ami Gates

University Of Colorado Boulder

Master of Science in Data Science

STAT 5000-002

Contents

1. Introduction	5
1.1 Data collection	6
1.1.1 COVID-19 Data	6
1.1.2 Food Production	8
1.2 Questions:	9
2. Analyses	10
2.1 Missing value treatment	10
2.2 Outlier treatment	11
2.2.1 COVID-19 Data: Column Cases_100K	11
2.2.2 COVID-19 Data: Column Deaths_100K	11
2.2.3 Food Production Data: Column Y2020	12
2.3 Exploratory Data analysis:	13
2.3.1 Dataset COVID-19	13
2.3.1.1 Univariate analysis for COVID-19 Dataset	13
2.3.1.2 Multivariate analysis on COVID-19 Dataset	15
2.3.2 Food Production Dataset	18
2.3.2.1 Univariate analysis for Food Production Dataset	18
2.3.2.2 Bivariate analysis for Food production dataset	19
2.3.2.3 Multivariate analysis for Food production dataset	20
2.4 Data Exploration	22
2.4.1 Relative Probability Distributions	22
2.4.1.1 Relative Probability Distribution for COVID-19 Dataset	22
2.4.1.2 Relative Probability Distribution for COVID-19 Dataset & Food Production Dataset	23
2.4.1.3 Relative Probability Distribution for AnnualTradeAnnouncementsTally	25
2.4.2. Measure of center and Variation- and Distributions	28
2.4.2.1 Measure of center and Variation for Food Production Dataset & COVID-19 Dataset	28
2.4.2.2 Measure of center and Variation for AnnualTradeAnnouncementsTally	32
3. Results:	38
3.1 Food Security Environment Scores 2017 & 2021	38
3.2 Trade Policies before and after COVID-19	43
3.3 Trade Policies announced globally	45
4. Conclusion:	46
5. References:	47

List of Figures:

Figure 1 Covid-19 Data before cleaning 1	6
Figure 2 COVID-19 Data before cleaning 2	7
Figure 3 COVID-19 Data after cleaning	7
Figure 4 Food Production Dataset 1	8
Figure 5 Food Production Dataset 2	8
Figure 6 Missing value detection in COVID-19 Data	10
Figure 7 Boxplot for column Cases_100K	11
Figure 8 Boxplot for Column Deaths_100K	12
Figure 9 Boxplot for Column Y2020	12
Figure 10 Summary of COVID-19 Dataset	13
Figure 11 Univariate analysis on COVID-19 Dataset	14
Figure 12 Multivariate plot of Column New_cases Vs. New_deaths Vs Year	15
Figure 13 Multivariate plot of Column New_cases Vs. Cumulative_cases Vs. Year	16
Figure 14 Multivariate plot of column Cumulative_deaths Vs. New_deaths Vs. Year	17
Figure 15 Univariate analysis on Food Production Dataset with 5-year increment	18
Figure 16 Bivariate analysis Y2000 Vs Item & Y2005 Vs. Item	19
Figure 17 Bivariate analysis for column Y2010 Vs. Item & Y2015 Vs. Item	19
Figure 18 Bivariate analysis between Y2020 Vs. Item	20
Figure 19 Multivariate analysis for columns Y2000 Vs. Y2005 Vs. Items & Y2010 Vs. Y2015 Vs. Items	20
Figure 20 Multivariate analysis for columns Y2019 Vs. Y2015 Vs. Items & Y2019 Vs. Y2020 Vs. Items	21
Figure 21 Histogram for column Cases_100k	22
Figure 22 Probability Density for Column Cases_100k	23
Figure 23 Histogram for Meat production per Population percentage	24
Figure 24 Probability Density for Column Meat production per Population percentage	24
Figure 25 Pie Chart for Number of Trade policies per country	25
Figure 26 Probability Density for column Number of import total	26
Figure 27 Probability for Column Number of Exports policies implemented	26
Figure 28 Probability for Column Number of Vaccine policies implemented in total	27
Figure 29 Histogram for column Item == Meat, Total	28
Figure 30 Box plot of meat produced by country	29
Figure 31 KDE for 2020 cases (Gaussian)	30
Figure 32 KDE for 2020 cases (Rectangular)	30
Figure 33 Q-Q plot for COVID cases in 2020	31
Figure 34 Histogram New Trade Policies Yearly	32
Figure 35 Histogram New Trade Policies Yearly (Logarithmic Scale)	33
Figure 36 Violin plot for New Trade policies by year	34
Figure 37 Violin Plots of New Trade Policies by year, Logarithmic Scale	34
Figure 38 Boxplot of New Trade Policies by Year, Logarithmic scale	35
Figure 39 KDF of New Trade Policies by year, Log scale, Triangular KDF	36

Figure 40 KDF of New Trade Policies by year, Log scale, epanechnikov KDF	36
Figure 41 Q-Q Plot for New Trade Policies per Country, grouped by year	37
Figure 42 Food Security Environment Scores 2017	38
Figure 43 Food Security Environment Scores 2021	39
Figure 44 Summaries of Year 2017 & 2021	40
Figure 45 Summary of Countries cases per 100K people from 2020	41
Figure 46 Change in Trade Policies Pre/Post Covid	45

1. Introduction

A pandemic is not a new event encountered in the history of humanity because mankind has faced various pandemics in history. The common point of pandemics is their serious negative effects on the global economy. Considering the food supply chain, one of the most important sectors of the economy, it has been seen that COVID-19 has an impact on the whole process from the field to the consumer. Considering, recent challenges in the food supply chain, there is now considerable concern about food production, processing, distribution, and demand. COVID-19 resulted in the movement restrictions of workers, changes in demand of consumers, closure of food production facilities, restricted food trade policies, and financial pressures in the food supply chain. Therefore, governments should facilitate the movement of workers and agri-food products.

The global disruption caused by COVID-19 has brought about several effects on the environment and climate. Due to movement restrictions and a significant slowdown of social and economic activities, air quality has improved in many cities with a reduction in water pollution in different parts of the world.

Supplies of staple crops are large, production prospects are favorable, and cereal stocks are expected to reach their third highest level on record. Moreover, most countries have designated the agriculture and agro-food sectors as essential and exempt from business closure and restrictions on movement. For many countries, the direct impacts of the pandemic on primary agriculture should be limited, as the disease does not affect the natural resources upon which production is based. However, the virus poses a serious threat to food security and livelihoods in the poorest countries, where agricultural production systems are more labor-intensive and there is less capacity to withstand a severe macroeconomic shock.

Because food is a necessity, the level of food demand should be affected less by the crisis than the demand for other goods and services. However, there has been a major shift in the structure of demand, with a collapse in demand from restaurants, hotels, and catering, the closure of open markets, and a surge in demand from supermarkets. There are signs that businesses along the food chain are already adapting to shifts in demand, for example by switching production lines and increasing their capacity to manage larger inventories, moving to online platforms and direct delivery to households, and hiring temporary staff. In all but the poorest countries, the biggest challenges for the sector come from the measures needed to contain COVID-19; the necessary adjustments within the sector to comply with those measures (which may increase costs); and the need to find alternative markets for products affected as people change their consumption habits in response to COVID-19.

Study starts with data collection and discusses the data collected and its significance, to make useful insights data cleaning is performed as missing value treatment and outlier detection and treatment. Next study discusses the questions which tend to solve like Did production change when covid hit? What about concentration on certain foods? Etc. Next section analysis helps to understand the data and data description more clearly. To understand and to perform test on data univariate, bi-variate, and multivariate value distribution is studied. Later, in data exploration we studied the relative probability distribution, measure of center and variances, using language R. Next section results we performed hypothesis testing on variables and challenge the significance

of Covid to population, income, and trade. At the end, we concluded the results based on results, visualizations, and distributions. Following statement sites, the study aim.

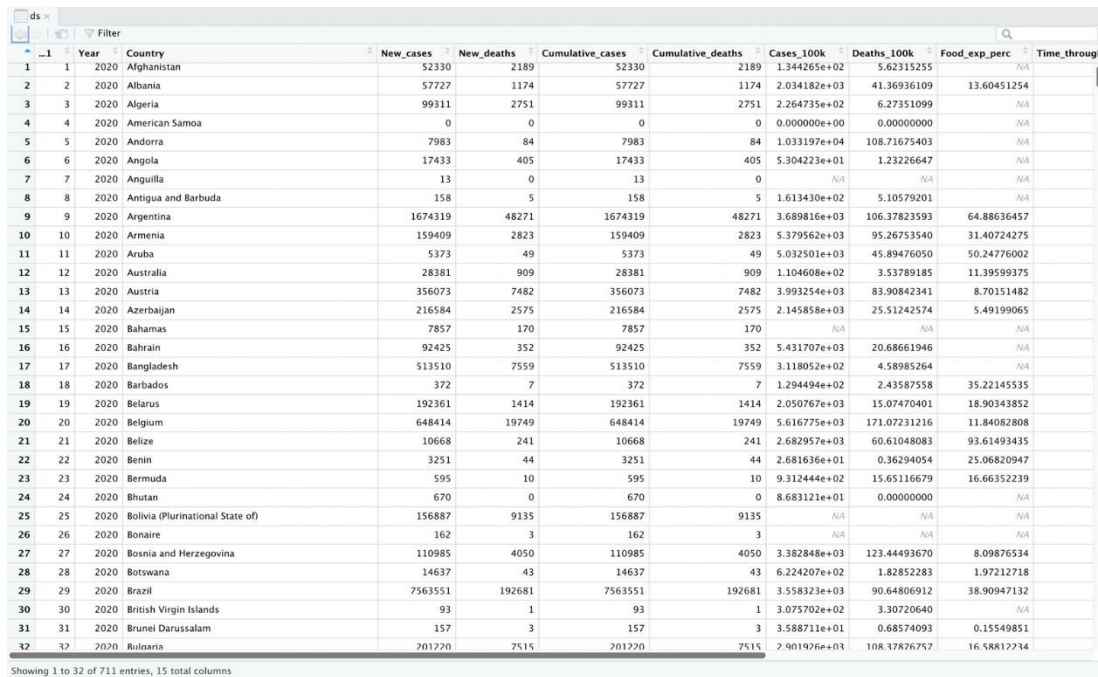
This study aims to gain insight into the Impacts of COVID-19 on food production resulting in human life around the world. Production of food in developed and developing countries compared to other countries. Also, COVID-19 took many lives and created complications for others.

1.1 Data collection

Data has been collected and stored in .csv (comma-separated values) from the sites like WHO (World Health Organization), and FAOSTAT (Food and Agriculture Organization Statistics).

1.1.1 COVID-19 Data

The COVID-19 dataset contains the values related to COVID-19 cases and deaths from 2020 to 2022. Countries are repeated with years from 2020 to 2022. This dataset is a combination of country-wise cases and food import-export. The dataset contains 711 records and 15 features.



Year	Country	New_cases	New_deaths	Cumulative_cases	Cumulative_deaths	Cases_100k	Deaths_100k	Food_exp_perc	Time_through
2020	Afghanistan	52330	2189	52330	2189	1.344265e+02	5.62315255	N/A	
2020	Albania	57727	1174	57727	1174	2.034182e+03	41.36936109	13.60451254	
2020	Algeria	99311	2751	99311	2751	2.264735e+02	6.27351099	N/A	
2020	American Samoa	0	0	0	0	0.000000e+00	0.00000000	N/A	
2020	Andorra	7983	84	7983	84	1.033197e+04	108.71675403	N/A	
2020	Angola	17433	405	17433	405	5.304223e+01	1.23226647	N/A	
2020	Anguilla	13	0	13	0	N/A	N/A	N/A	
2020	Antigua and Barbuda	158	5	158	5	1.613430e+02	5.10579201	N/A	
2020	Argentina	1674319	48271	1674319	48271	3.689816e+03	106.37823593	64.88636457	
2020	Armenia	159409	2823	159409	2823	5.379562e+03	95.26753540	31.40724275	
2020	Aruba	5373	49	5373	49	5.032501e+03	45.89476050	50.24776002	
2020	Australia	28381	909	28381	909	1.104608e+02	3.53789185	11.39599375	
2020	Austria	356073	7482	356073	7482	3.993254e+03	83.90842341	8.70151482	
2020	Azerbaijan	216584	2575	216584	2575	2.145858e+03	25.51242574	5.49199065	
2020	Bahamas	7857	170	7857	170	N/A	N/A	N/A	
2020	Bahrain	92425	352	92425	352	5.431707e+03	20.68661946	N/A	
2020	Bangladesh	513510	7559	513510	7559	3.118052e+02	4.58985264	N/A	
2020	Barbados	372	7	372	7	1.294494e+02	2.43587558	35.22145535	
2020	Belarus	192361	1414	192361	1414	2.050767e+03	15.07470401	18.90343852	
2020	Belgium	648414	19749	648414	19749	5.616775e+03	171.07231216	11.84082808	
2020	Belize	10668	241	10668	241	2.682957e+03	60.61048083	93.61493435	
2020	Benin	3251	44	3251	44	2.681636e+01	0.36294054	25.06820947	
2020	Bermuda	595	10	595	10	9.312444e+02	15.65116679	16.66352239	
2020	Bhutan	670	0	670	0	8.683121e+01	0.00000000	N/A	
2020	Bolivia (Plurinational State of)	156887	9135	156887	9135	N/A	N/A	N/A	
2020	Bonaire	162	3	162	3	N/A	N/A	N/A	
2020	Bosnia and Herzegovina	110985	4050	110985	4050	3.382848e+03	123.44493670	8.09876534	
2020	Botswana	14637	43	14637	43	6.224207e+02	1.82852283	1.97212718	
2020	Brazil	7563551	192681	7563551	192681	3.558323e+03	90.64806912	38.90947132	
2020	British Virgin Islands	93	1	93	1	3.075702e+02	3.30720640	N/A	
2020	Brunei Darussalam	157	3	157	3	3.588711e+01	0.68574093	0.15549851	
2020	Ruoinaria	201220	7515	201220	7515	2.901926e+03	108.37876757	16.58812234	

Figure 1 Covid-19 Data before cleaning 1

- Data cleaning activity is performed on dataset like missing value imputation, outlier treatment and checking for value distribution in the columns.
- After data cleaning procedures data is used to visualize each column for its distribution.

- Boxplots, Histograms & scatter plots are used in terms of the univariate and multivariate analysis.

New_cases	New_deaths	Cumulative_cases	Cumulative_deaths	Cases_100k	Deaths_100k	Food_exp_perc	Time_through_customs	Agri_exp_perc	Food_imp_perc	Food_Production
52330	2189	52330	2189	1.344265e+02	5.62315255	NA	NA	NA	NA	NA
57727	1174	57727	1174	2.034182e+03	41.36936109	13.60451254	NA	0.565251251	16.367330	NA
99311	2751	99311	2751	2.264735e+02	6.27351099	NA	NA	NA	NA	NA
0	0	0	0	0.000000e+00	0.00000000	NA	NA	NA	NA	NA
7983	84	7983	84	1.033197e+04	108.71675403	NA	NA	NA	NA	NA
17433	405	17433	405	5.304223e+01	1.23226647	NA	NA	NA	NA	NA
13	0	13	0	NA	NA	NA	NA	NA	NA	NA
158	5	158	5	1.613430e+02	5.10579201	NA	NA	NA	NA	NA
1674319	48271	1674319	48271	3.689816e+03	106.37823593	64.88636457	NA	0.875916610	8.765561	NA
159409	2823	159409	2823	5.379562e+03	95.26753540	31.40724275	12.0	0.324368741	17.713523	NA
5373	49	5373	49	5.032501e+03	45.89476050	50.24776002	NA	0.069199439	30.610421	NA
28381	909	28381	909	1.104608e+02	3.53789185	11.39599375	NA	1.429036118	7.607558	NA
356073	7482	356073	7482	3.993254e+03	83.90842341	8.70151482	NA	1.519470540	8.402001	NA
216584	2575	216584	2575	2.145858e+03	25.51242574	5.49199065	NA	1.029853904	17.331727	NA
7857	170	7857	170	NA	NA	NA	NA	NA	NA	NA
92425	352	92425	352	5.431707e+03	20.68661946	NA	NA	NA	NA	NA
513510	7559	513510	7559	3.118052e+02	4.58985264	NA	NA	NA	NA	NA
372	7	372	7	1.294494e+02	2.43587558	35.22145535	NA	0.023730130	23.082624	NA
192361	1414	192361	1414	2.050767e+03	15.07470401	18.90343852	NA	4.010628613	11.370028	NA
648414	19749	648414	19749	5.616775e+03	171.07231216	11.84082808	6.1	1.219520263	10.778694	NA
10668	241	10668	241	2.682957e+03	60.61048083	93.61493435	NA	0.737496381	25.648716	NA
3251	44	3251	44	2.681636e+01	0.36294054	25.06820947	NA	64.643800570	32.821735	NA
595	10	595	10	9.312444e+02	15.65116679	16.66352239	NA	0.000581045	23.832469	NA
670	0	670	0	8.683121e+01	0.00000000	NA	NA	NA	NA	NA
156887	9135	156887	9135	NA	NA	NA	NA	NA	NA	NA
162	3	162	3	NA	NA	NA	NA	NA	NA	NA
110985	4050	110985	4050	3.382848e+03	123.44493670	8.09876534	NA	4.945927881	17.820751	NA
14637	43	14637	43	6.224207e+02	1.82852283	1.97212718	NA	0.097379886	12.709408	NA
7563551	192681	7563551	192681	3.558323e+03	90.64806912	38.90947132	NA	5.662895667	6.300103	NA
93	1	93	1	3.075702e+02	3.30720640	NA	NA	NA	NA	NA
157	3	157	3	3.588711e+01	0.68574093	0.15549851	NA	0.018895330	10.766720	NA

Showing 1 to 31 of 711 entries, 15 total columns

Figure 2 COVID-19 Data before cleaning 2

Year	Country	New_cases	New_deaths	Cumulative_cases	Cumulative_deaths	Cases_100k	Deaths_100k
1	2020 Afghanistan	52330	2189	52330	2189	1.344265e+02	5.62315255
2	2020 Albania	57727	1174	57727	1174	2.034182e+03	41.36936109
3	2020 Algeria	99311	2751	99311	2751	2.264735e+02	6.27351099
4	2020 American Samoa	0	0	0	0	0.000000e+00	0.00000000
5	2020 Andorra	7983	84	7983	84	1.033197e+04	108.71675403
6	2020 Angola	17433	405	17433	405	5.304223e+01	1.23226647
7	2020 Anguilla	13	0	13	0	2.109835e+03	16.73112423
8	2020 Antigua and Barbuda	158	5	158	5	1.613430e+02	5.10579201
9	2020 Argentina	1674319	48271	1674319	48271	3.689816e+03	106.37823593
10	2020 Armenia	159409	2823	159409	2823	5.379562e+03	95.26753540
11	2020 Aruba	5373	49	5373	49	5.032501e+03	45.89476050
12	2020 Australia	28381	909	28381	909	1.104608e+02	3.53789185
13	2020 Austria	356073	7482	356073	7482	3.993254e+03	83.90842341
14	2020 Azerbaijan	216584	2575	216584	2575	2.145858e+03	25.51242574
15	2020 Bahamas	7857	170	7857	170	2.109835e+03	16.73112423
16	2020 Bahrain	92425	352	92425	352	5.431707e+03	20.68661946
17	2020 Bangladesh	513510	7559	513510	7559	3.118052e+02	4.58985264
18	2020 Barbados	372	7	372	7	1.294494e+02	2.43587558
19	2020 Belarus	192361	1414	192361	1414	2.050767e+03	15.07470401
20	2020 Belgium	648414	19749	648414	19749	5.616775e+03	171.07231216
21	2020 Belize	10668	241	10668	241	2.682957e+03	60.61048083
22	2020 Benin	3251	44	3251	44	2.681636e+01	0.36294054
23	2020 Bermuda	595	10	595	10	9.312444e+02	15.65116679
24	2020 Bhutan	670	0	670	0	8.683121e+01	0.00000000
25	2020 Bolivia (Plurinational State of)	156887	9135	156887	9135	2.109835e+03	16.73112423
26	2020 Bonaire	162	3	162	3	2.109835e+03	16.73112423
27	2020 Bosnia and Herzegovina	110985	4050	110985	4050	3.382848e+03	123.44493670
28	2020 Botswana	14637	43	14637	43	6.224207e+02	1.82852283
29	2020 Brazil	7563551	192681	7563551	192681	3.558323e+03	90.64806912
30	2020 British Virgin Islands	93	1	93	1	3.075702e+02	3.30720640
31	2020 Brunei Darussalam	157	3	157	3	3.588711e+01	0.68574093
32	2020 Bulgaria	201220	7515	201220	7515	2.901926e+03	108.37876757

Showing 1 to 32 of 711 entries, 8 total columns

Figure 3 COVID-19 Data after cleaning

1.1.2 Food Production

One of the datasets contains the food and agriculture organization of the United Nations. The dataset contains countries and their production statistics from the year 2000 to the year 2020. Crop and livestock statistics are recorded for 278 products, covering the following categories:

- 1) CROP'S PRIMARY
- 2) CROPS PROCESSED
- 3) LIVE ANIMALS
- 4) LIVESTOCK PRIMARY
- 5) LIVESTOCK PROCESSED

This dataset contains 2227 records and 24 features. This dataset does not contain any missing value.

...1	Area	Item	Y2000	Y2001	Y2002	Y2003	Y2004	Y2005	Y2006	Y2007	Y2008
1	Afghanistan	Cereals, Total	1940000	2108000	3737000	4381000	3560000	5425000	4638000	5786000	3950000
2	Afghanistan	Citrus Fruit, Total	17794	17760	17721	21065	16931	15200	10284	8709	8100
3	Afghanistan	Eggs Primary	15010	13100	16730	16371	17400	21824	14688	16858	16698
4	Afghanistan	Fruit Primary	856631	894887	917769	969753	801156	936922	1241884	1276450	1244515
5	Afghanistan	Meat, Total	320785	278743	304321	298619	322977	318615	272400	267240	294060
6	Afghanistan	Milk, Total	1651327	1106152	1862800	1734386	1792274	1737674	1640689	1756000	1707200
7	Afghanistan	Pulses, Total	47015	47329	47585	42100	42000	47642	42100	40505	51000
8	Afghanistan	Roots and Tubers, Total	235000	235000	230000	350000	300000	300000	300000	300800	280000
9	Afghanistan	Sugar Crops Primary	39000	39000	39000	61000	65000	41000	73000	86938	76276
10	Afghanistan	Treenuts, Total	20800	23800	20574	20900	21945	23335	28350	44387	55833
11	Afghanistan	Vegetables Primary	1108694	1093772	1083073	1065317	1070651	1243836	1037022	1040207	1009392
12	Albania	Cereals, Total	565800	502500	518800	488500	498800	511200	507500	493600	608500
13	Albania	Citrus Fruit, Total	2600	2800	3400	4400	6262	6800	5600	6725	8640
14	Albania	Eggs Primary	21490	24550	25050	27200	29500	27750	26850	27500	29970
15	Albania	Fruit Primary	391651	453804	459800	426800	431002	455290	462908	516200	527548

Figure 4 Food Production Dataset 1

Y2009	Y2010	Y2011	Y2012	Y2013	Y2014	Y2015	Y2016	Y2017	Y2018	Y2019	Y2020
6514000	5957000	4681020	6379000	6520329	6748023	5808288	5532695	4894365	4134191	5583461	6025977
6000	4000	8000	8000	9320	9930	8028	22049	20558	21811	37262	6776
16778	16344	17210	18000	18000	18500	19833	19884	20000	19906	19930	19945
1459939	1360110	1274113	1352528	1346187	1715054	1831103	2597914	2870795	2290670	3112970	3459051
304340	328160	335664	330600	322110	311020	302780	295536	311953	312780	268273	312773
1732000	1721000	1743000	1844700	1835200	2418764	2177252	2168253	2099287	2153910	2111786	2137132
50000	53300	67000	69000	60000	60000	106651	81053	36304	41002	43645	43556
302400	246000	205000	230000	302980	340257	327507	427917	513194	615684	921122	855395
139700	107600	107696	91564	104645	86412	50949	18946	35827	30784	45822	45324
56365	71463	78312	79831	57217	41885	32288	43487	38649	49461	57094	58414
1022578	969763	762948	817744	790267	973722	1008465	1466532	922563	887174	1400926	1822632
630900	693800	701200	697400	702870	700370	695000	698430	701734	678196	666065	684023
10200	13505	15000	18860	19996	21357	29999	40339	41032	45500	46882	49201
30250	31500	31945	50759	47510	47807	52487	53150	51299	52333	54727	54444
565754	634378	663653	703600	692054	708365	765329	810483	805534	786761	809337	818946

Figure 5 Food Production Dataset 2

- Figure 4 & Figure 5 shows the Food production values from the year 2000 to year 2020 for various countries in terms of crops and livestock.

1.2 Questions:

To Answer following questions this study is conducted. If needs more data can be added.

- 1) Did production change when covid hit?
- 2) What about concentration on certain foods?
- 3) What about certain countries? did some get hit harder than others?
- 4) Did certain countries start exporting more when covid hit?
- 5) Did certain countries start importing more when covid hit?
- 6) How does import/export rates of food affect a country's food insecurity
- 7) Does covid modulate these aspects?
- 8) Does production improve with time?
- 9) Do we always remove outliers from the dataset?
- 10) Does this data will be sufficient to answers mentioned above?

2. Analyses

2.1 Missing value treatment

- In this dataset, many columns contain missing values like column "Food_exp_perc", "Time_through_customs", "Agri_exp_perc", "Food_imp_perc", "Food_Production", "Food_bev_vat". These columns contain missing values ranging from Approx. 72% to 98%. Hence, decided to drop these columns. Now, in COVID-19 Dataset only 8 features left.
- Next part of cleaning process is imputing missing values in columns which have missing values less than 50%. We have two columns i.e., Cases_100K & Deaths_100K having 24% missing values. When we check the distribution of values for column "Cases_100K", it's shows that mean is greater than median values which means that values are negatively distributed/ negatively skewed.
- Distribution of values present in column "Deaths_100K" is same as column "Cases_100K" with mean values greater than median values which means values are negatively skewed.
- Considering, details of value distributions in columns "Cases_100K" and "Deaths_100K". Imputing missing values with median value is considered correct. Hence missing values of respective columns are replaced with median values.

```
> colSums(is.na(ds))
```

...	1	Year	Country	New_cases	New_deaths
	0	0	0	0	0
Cumulative_cases	Cumulative_deaths	Cases_100k	Deaths_100k	Food_exp_perc	
	0	0	173	173	513
Time_through_customs	Agri_exp_perc	Food_imp_perc	Food_Production	Food_bev_vat	
	701	514	513	711	642

```
> |
```

Figure 6 Missing value detection in COVID-19 Data

2.2 Outlier treatment

2.2.1 COVID-19 Data: Column Cases_100K

- Next, Outlier treatment of columns. Plotting boxplot helps to understand the spread of values in distribution. Hence, performed the outlier treatment on columns having outliers. From Figure 5, Boxplot for column Cases_100K, this plot shows the outlier values in the column.
- Column Cases_100K shows the new cases detected per 100K testing.

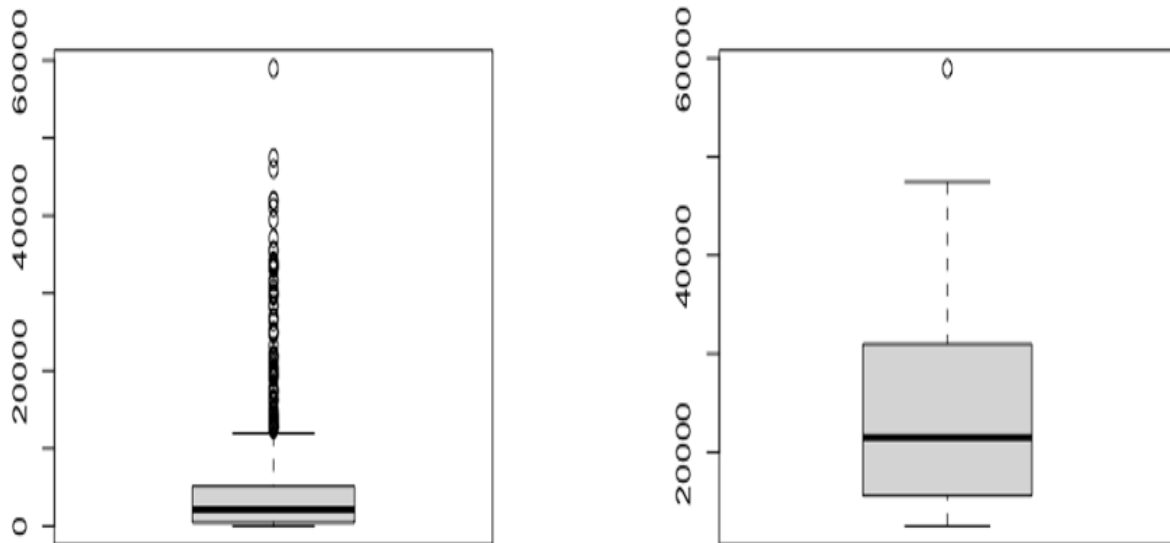


Figure 7 Boxplot for column Cases_100K

2.2.2 COVID-19 Data: Column Deaths_100K

- Column Deaths_100K shows the deaths per 100K testing. Boxplot shows the spread of values in the column distributions.

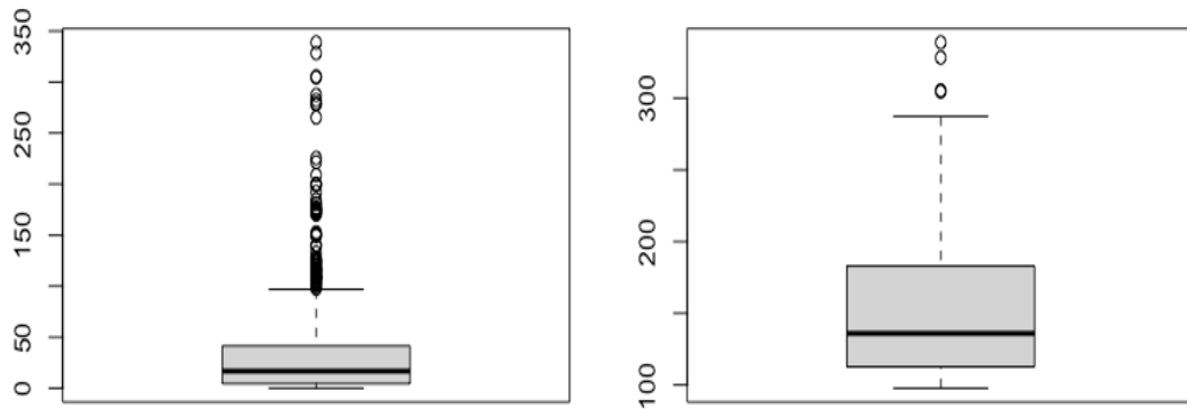


Figure 8 Boxplot for Column Deaths_100K

- Like, Columns Cases_100K & Deaths_100K we have capped the values of columns in dataset COVID-19.
- For Dataset Food production, shows outlier values but, removing those value can cause information loss from countries whose food production increased for year due to good conditions.

2.2.3 Food Production Data: Column Y2020

- Column Y2020 shows distribution of values for Year 2020. It also shows the values for all the countries whose production in various items Crops, Livestock.

For the Year 2020

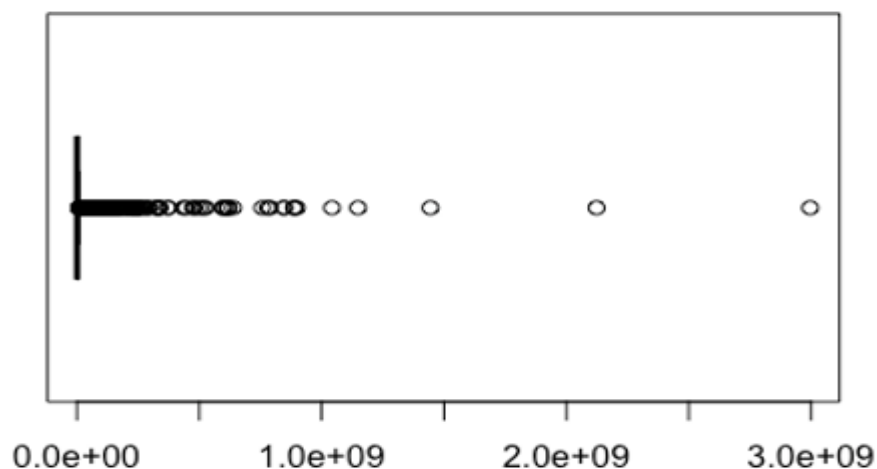


Figure 9 Boxplot for Column Y2020

2.3 Exploratory Data analysis:

2.3.1 Dataset COVID-19

- In Exploratory data analysis (EDA) columns of dataset are examined to check its distribution, its variation considering other columns i.e., its correlation with other columns.
- In R, we can check the value distributions and other attributes using summary command as shown in figure 8 Summary of COVID-19 Dataset.

```

...1      Year      Country      New_cases      New_deaths      Cumulative_cases
Min.   : 1.0   Min.   :2020   Length:711   Min.    :    0   Min.    :    0   Min.    :    0
1st Qu.:178.5 1st Qu.:2020   Class :character 1st Qu. :  4926 1st Qu. :   32 1st Qu. :  7314
Median :356.0 Median :2021   Mode  :character Median  : 36426 Median :   324 Median : 66626
Mean   :356.0 Mean   :2021           Mean   : 842576 Mean   :  9096 Mean   :1363073
3rd Qu.:533.5 3rd Qu.:2022           3rd Qu. :334232 3rd Qu. : 3340 3rd Qu. : 536543
Max.   :711.0 Max.   :2022           Max.   :39404260 Max.   :467012 Max.   :92953690

Cumulative_deaths  Cases_100k  Deaths_100k  Food_exp_perc  Time_through_customs  Agri_exp_perc
Min.   :    0   Min.   :  0.0   Min.   :  0.00   Min.   :  0.0568   Min.   :  2.900   Min.   :  0.0000
1st Qu.:   70   1st Qu.: 177.7   1st Qu.:  1.78   1st Qu.:  8.6064   1st Qu.:  3.425   1st Qu.:  0.6127
Median :   827   Median : 2109.8   Median : 16.73   Median :16.6065   Median :  6.100   Median :  1.2630
Mean   : 19474   Mean   : 5664.0   Mean   : 40.54   Mean   :24.7656   Mean   :  5.880   Mean   :  3.2161
3rd Qu.:  7498   3rd Qu.: 6584.8   3rd Qu.: 59.28   3rd Qu.:32.1702   3rd Qu.:  6.625   3rd Qu.:  3.5810
Max.   :1033207   Max.   :58957.1   Max.   :338.81   Max.   :95.4371   Max.   :12.000   Max.   :74.8808
NA's   :173      NA's   :173      NA's   :513      NA's   :701      NA's   :514

Food_imp_perc  Food_Production  Food_bev_vat
Min.   : 4.111   Mode:logical   Min.   : 3.376
1st Qu.:10.361   NA's:711       1st Qu.:13.144
Median :13.248           Median :19.521
Mean   :14.831           Mean   :23.207
3rd Qu.:17.769           3rd Qu.:31.342
Max.   :44.259           Max.   :62.685
NA's   :513              NA's   :642

```

Figure 10 Summary of COVID-19 Dataset

2.3.1.1 Univariate analysis for COVID-19 Dataset

- Quantitative variables are considered here, as only categorical columns are Year and Country. Hence, plotting histograms for cleaned data, with six columns in figure 9 Univariate analysis on COVID-19 dataset.

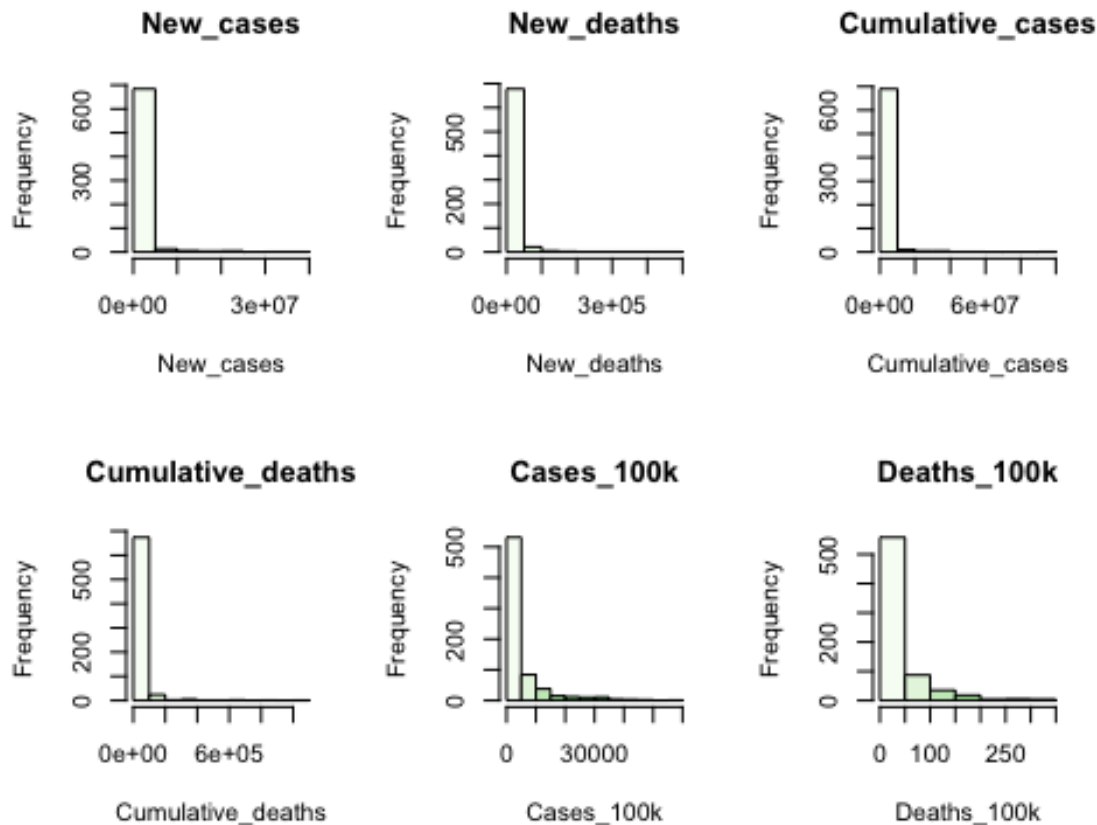


Figure 11 Univariate analysis on COVID-19 Dataset

- Column **New_cases**: Here, most values are close to zero with values in positive skewed distribution. Maximum frequency of New_cases go till 600 in frequency.
- Column **New_deaths**: This column explains the new confirmed cases by subtracting previous cumulative case count from current cumulative cases count. Histogram shows the value distribution close towards the zero.
- Column **Cumulative_cases**: This column explains Cumulative confirmed cases reported to WHO to date. Here, we can see the rise in the cases with low frequency and cases are reported close to the less count shows high frequency. It means that less count of cases reported more.
- Column **Cumulative_deaths**: This column explains Cumulative confirmed deaths reported to WHO to date. Plot shows the distribution of value close to the lower values with frequency more than 500. This graph explains that less deaths are reported maximum times.
- Column **Cases_100K**: This column explains Cumulative confirmed cases reported to WHO to date per 100,000 population. The value distribution shows values close to lower

count reported maximum times. It means that cases till 20K reported maximum times than 30K.

- Column **Deaths_100K**: This column explains the Cumulative confirmed deaths reported to WHO to date per 100,000 population. Plot for this column shows distribution of values close towards the lower values till 80. It means death count till 80 reported maximum times.

2.3.1.2 Multivariate analysis on COVID-19 Dataset

- Figure 10, Multivariate plot of column New_cases Vs. New_deaths Vs. Year shows the variation and distribution of values. From plot we can visualize that with maximum cases and deaths are happened in year 2020. Later we can see some outliers in year 2021 showing more cases with more deaths.

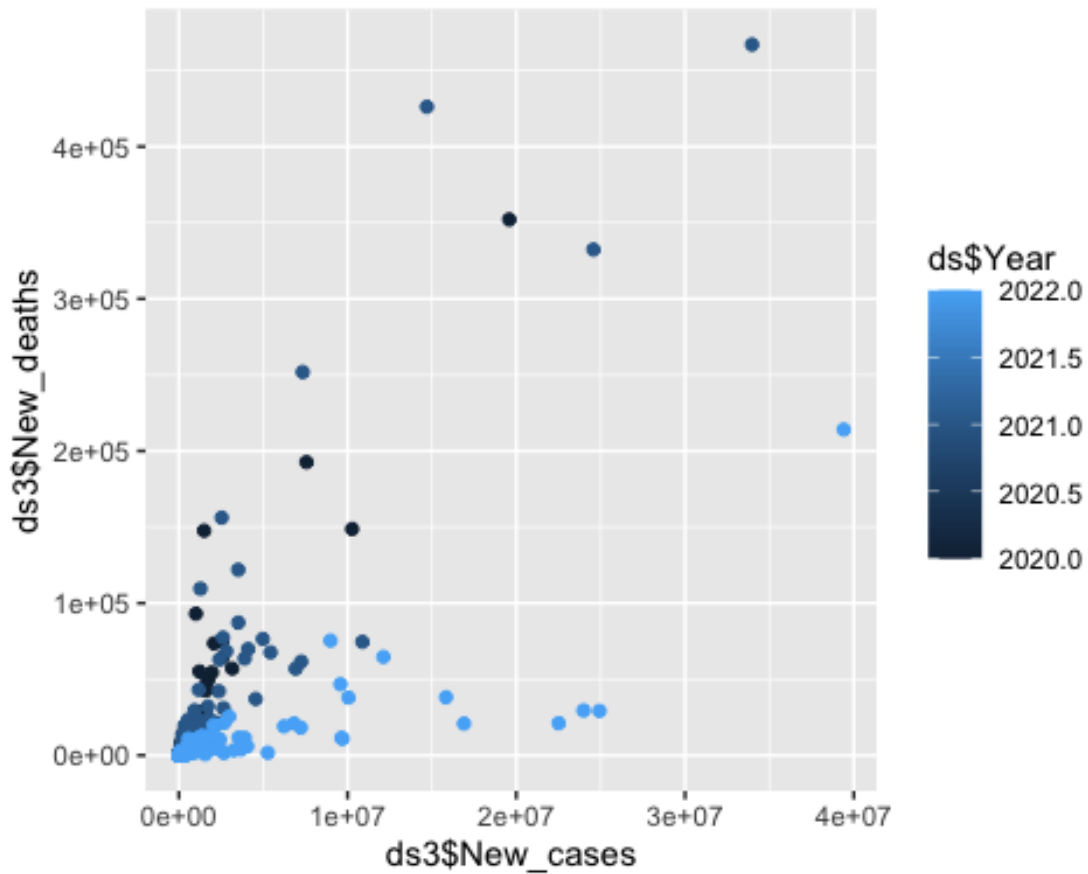


Figure 12 Multivariate plot of Column New_cases Vs. New_deaths Vs Year

- Figure 11 Multivariate plot of Column New_cases Vs. Cumulative_cases Vs. Year. Here plot shows the linear relationship between New_cases with Cumulative_cases, it is clear from plot as cumulative cases are high in year 2022. Also, dark points are from early time like 2020 showing high New_cases.

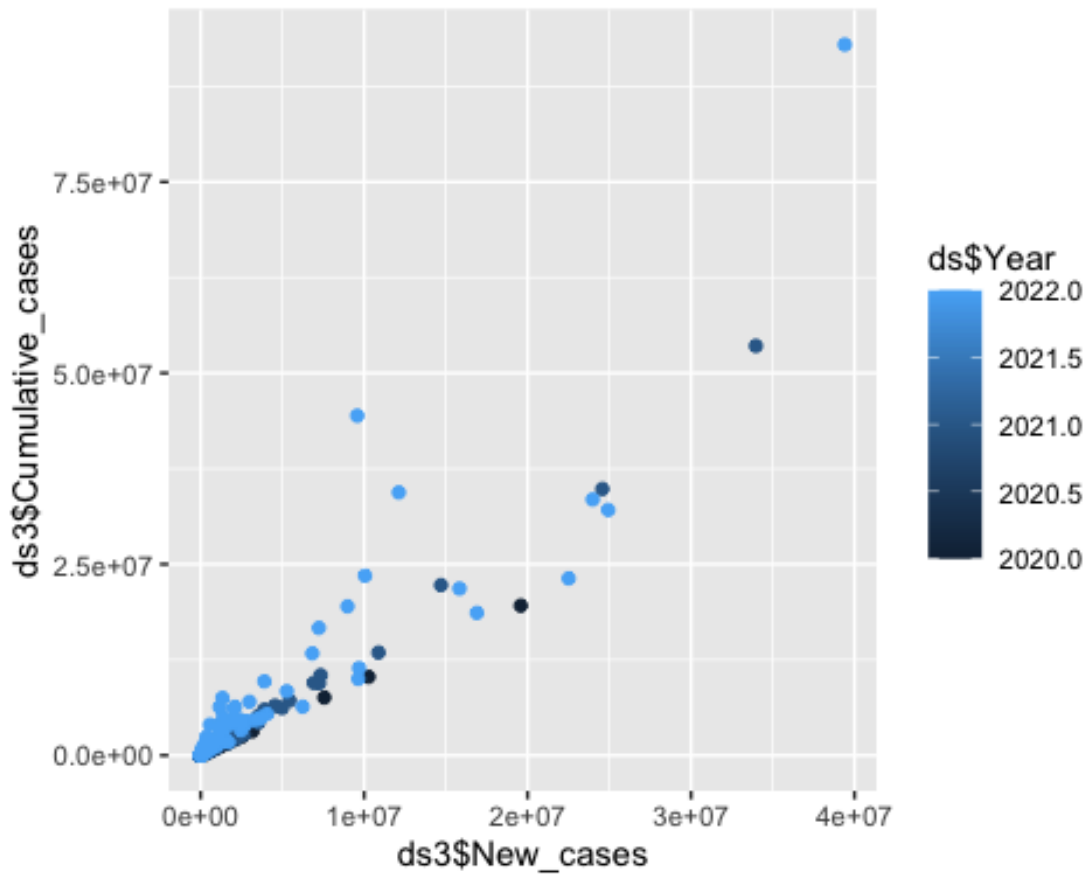


Figure 13 Multivariate plot of Column New_cases Vs. Cumulative_cases Vs. Year

- Next, from figure 12 Multivariate plot of column Cumulative_deaths Vs. New_deaths Vs. Year. Here plot shows linear relationship between Cumulative_deaths and New_deaths. From intensity of colors, we can conclude that maximum new deaths are in year 2020 and maximum cumulative deaths are in year 2022.

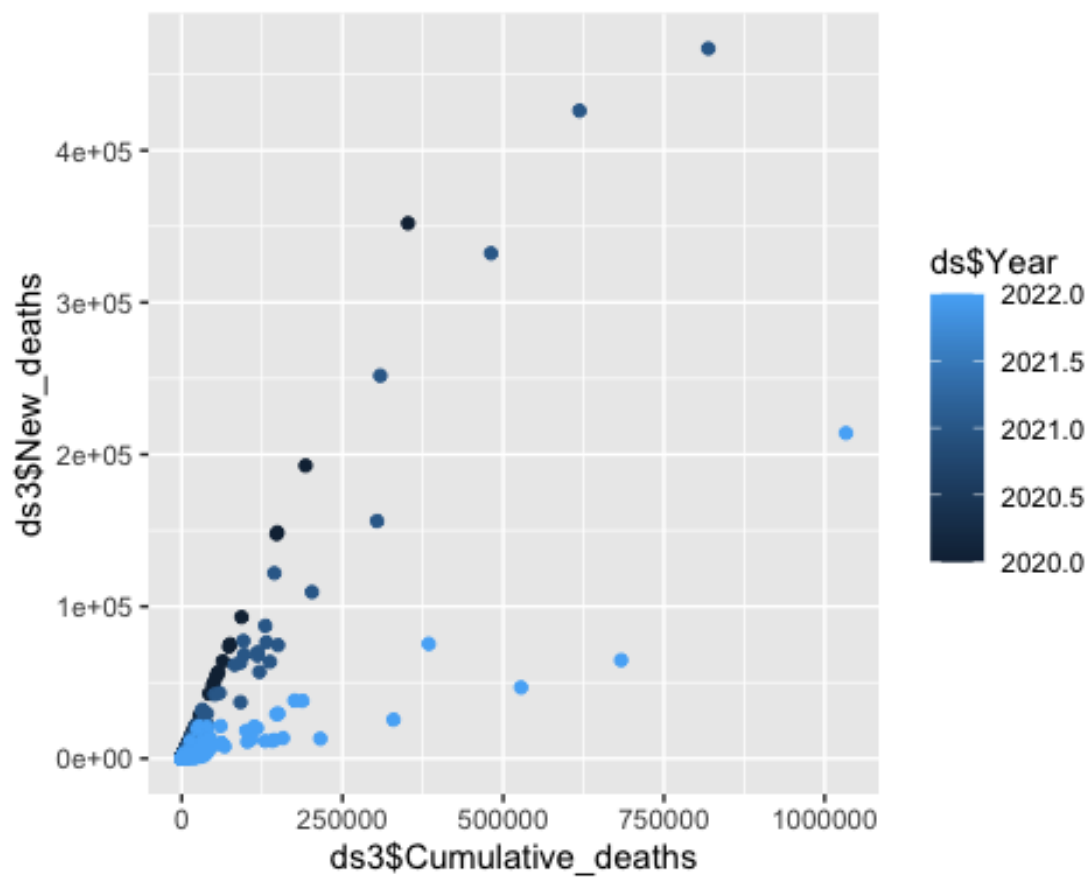


Figure 14 Multivariate plot of column Cumulative_deaths Vs. New_deaths Vs. Year

2.3.2 Food Production Dataset

Food production dataset contains the food and agriculture organization of the United Nations. The dataset contains countries and their production statistics from the year 2000 to the year 2020.

2.3.2.1 Univariate analysis for Food Production Dataset

- Crop and livestock statistics are recorded for 278 products dataset contains 2227 records and 24 features. This dataset does not contain any missing value.

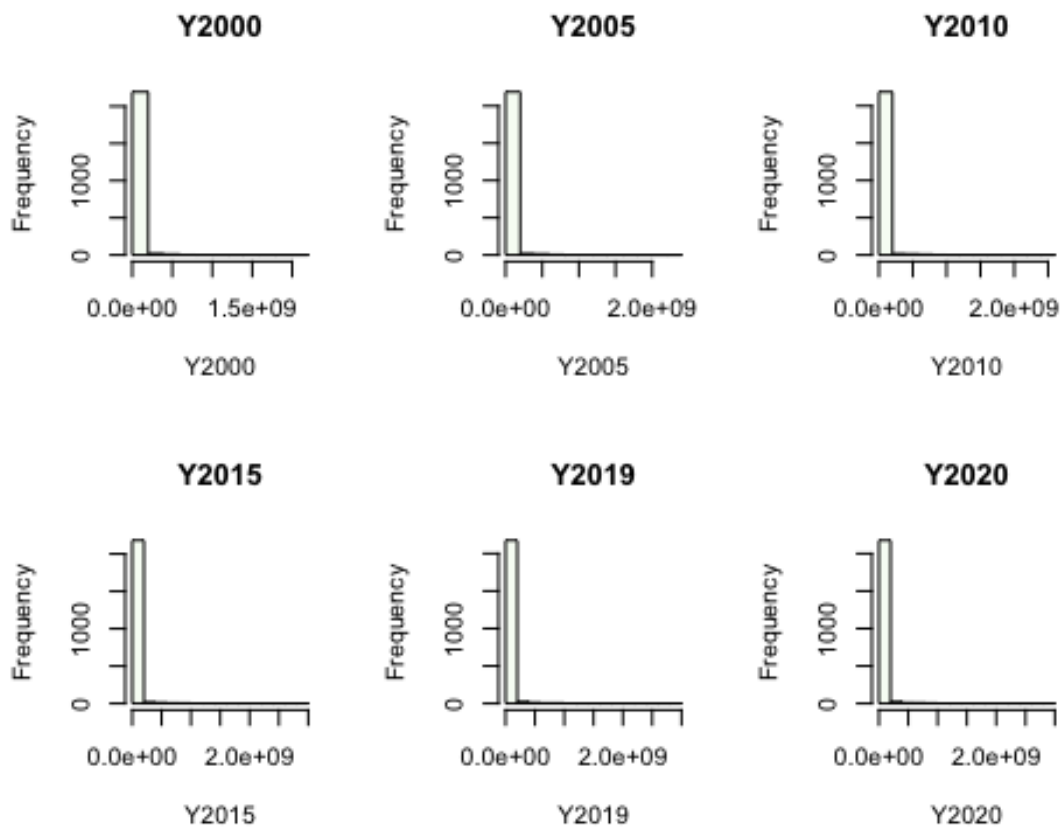


Figure 15 Univariate analysis on Food Production Dataset with 5-year increment

- From figure 15 univariate analysis on food production dataset with 5-year increment, graphs we can see that production is improved with incrementing years. Reason been all plots looks same as this dataset contains crops and livestock values with country wise fashion, so values are repeating in each column hence, all plots look similar.

2.3.2.2 Bivariate analysis for Food production dataset

- Figure 16 Bivariate analysis Y2000 Vs Items shows the amount of item produced in Year 2000. Similarly in second part of figure, plots show items produced in year 2005.

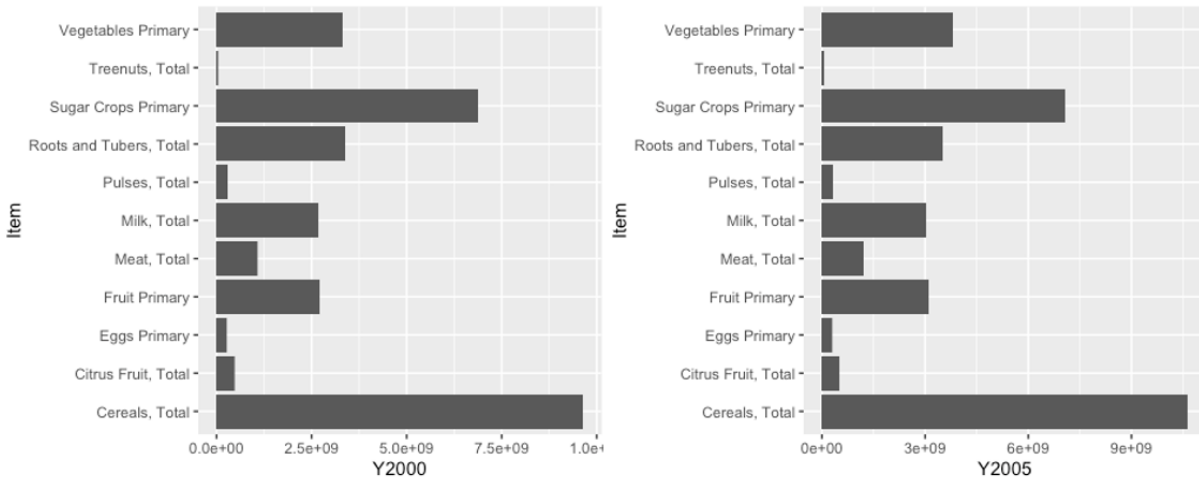


Figure 16 Bivariate analysis Y2000 Vs Item & Y2005 Vs. Item

- Figure 17, Bivariate analysis for column Y2010 Vs. Item, shows the production of the amount of item produced in Year 2010. Similarly in second part of figure, plots show items produced in year 2015.

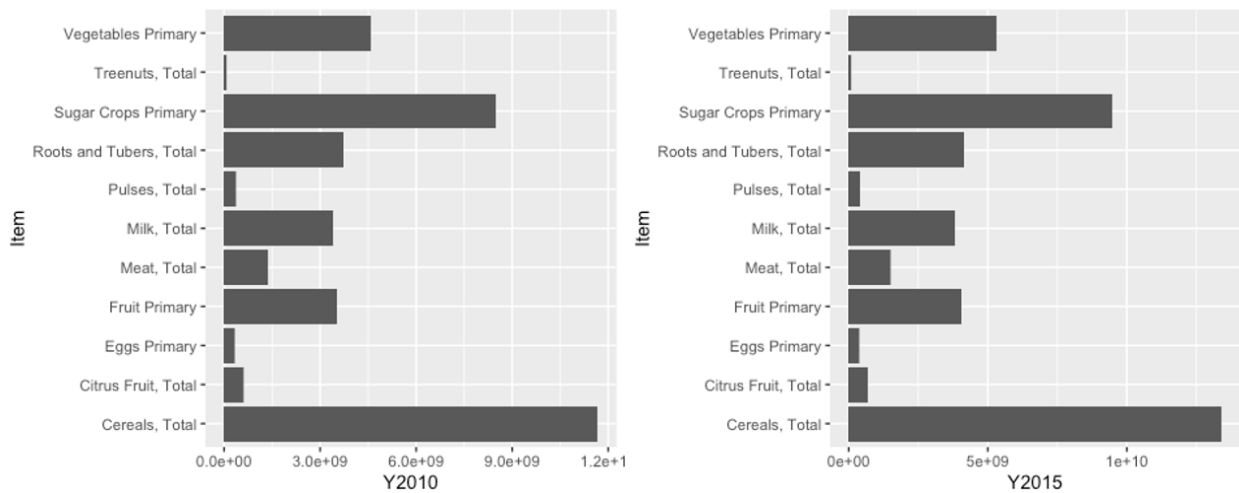


Figure 17 Bivariate analysis for column Y2010 Vs. Item & Y2015 Vs. Item

- In figure 18 Bivariate analysis between Y2020 Vs. Item show the amount of item produced in Year 2020.

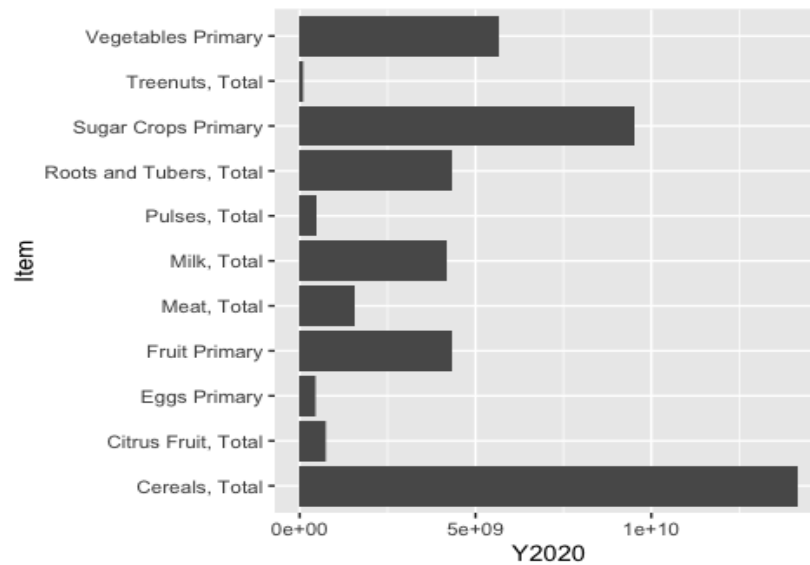


Figure 18 Bivariate analysis between Y2020 Vs. Item

2.3.2.3 Multivariate analysis for Food production dataset

- From figure 19, Multivariate analysis, Year column have linear relationship with other year columns and from item column have maximum production of Cereals, which can be verified from univariate analysis.

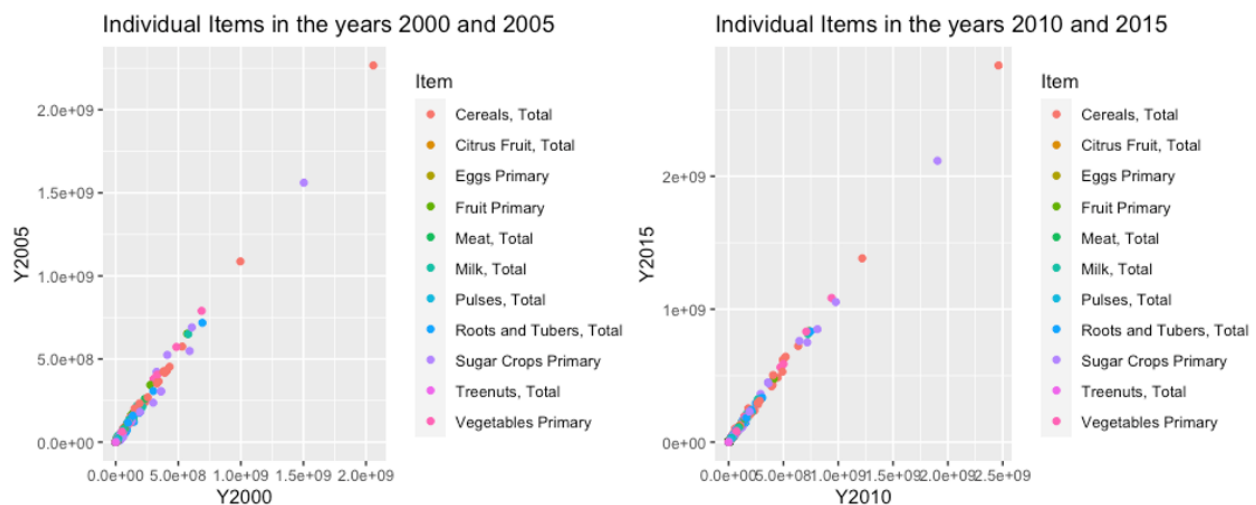


Figure 19 Multivariate analysis for columns Y2000 Vs. Y2005 Vs. Items & Y2010 Vs. Y2015 Vs. Items

- From figure 20, Multivariate analysis, Year column have linear relationship with other year columns, but with less variance in production. Compared to the other year data, linear relationship is improved.

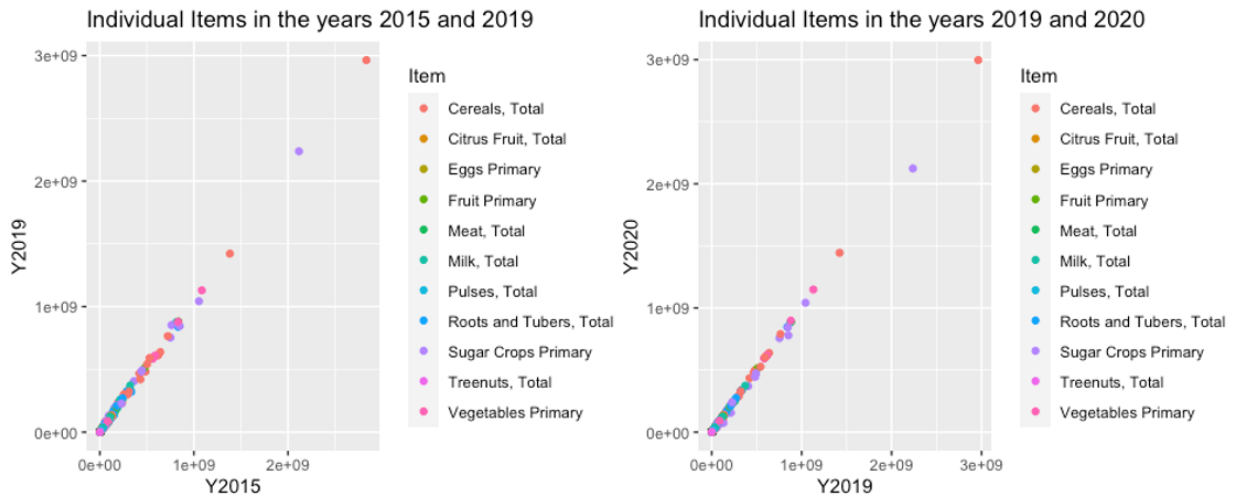


Figure 20 Multivariate analysis for columns Y2019 Vs. Y2015 Vs. Items & Y2019 Vs. Y2020 Vs. Items

2.4 Data Exploration

2.4.1 Relative Probability Distributions

A relative frequency distribution shows the proportion of the total number of observations associated with each value or class of values and is related to a probability distribution, which is extensively used in statistics. (R.H.Riffenburgh, 2012)

2.4.1.1 Relative Probability Distribution for COVID-19 Dataset

1. **Dataset:** COVID19_data_cleaned
2. **Column:** Cases_100k
3. **Maximum value :** 58957.14
4. **Minimum value :** 0
5. **Probability Density function:**

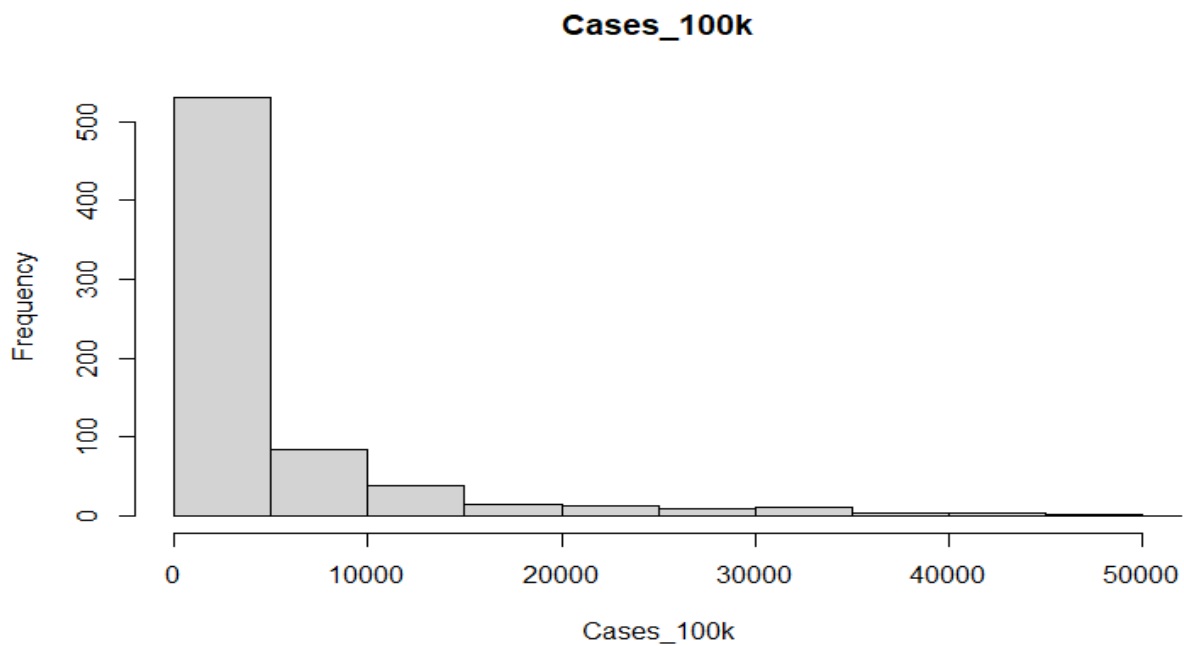


Figure 21 Histogram for column Cases_100k

- Inferences:
 - The number of cases_100k shows positively skewed, geometric distribution.

6. Probability Density for Column Cases_100k:

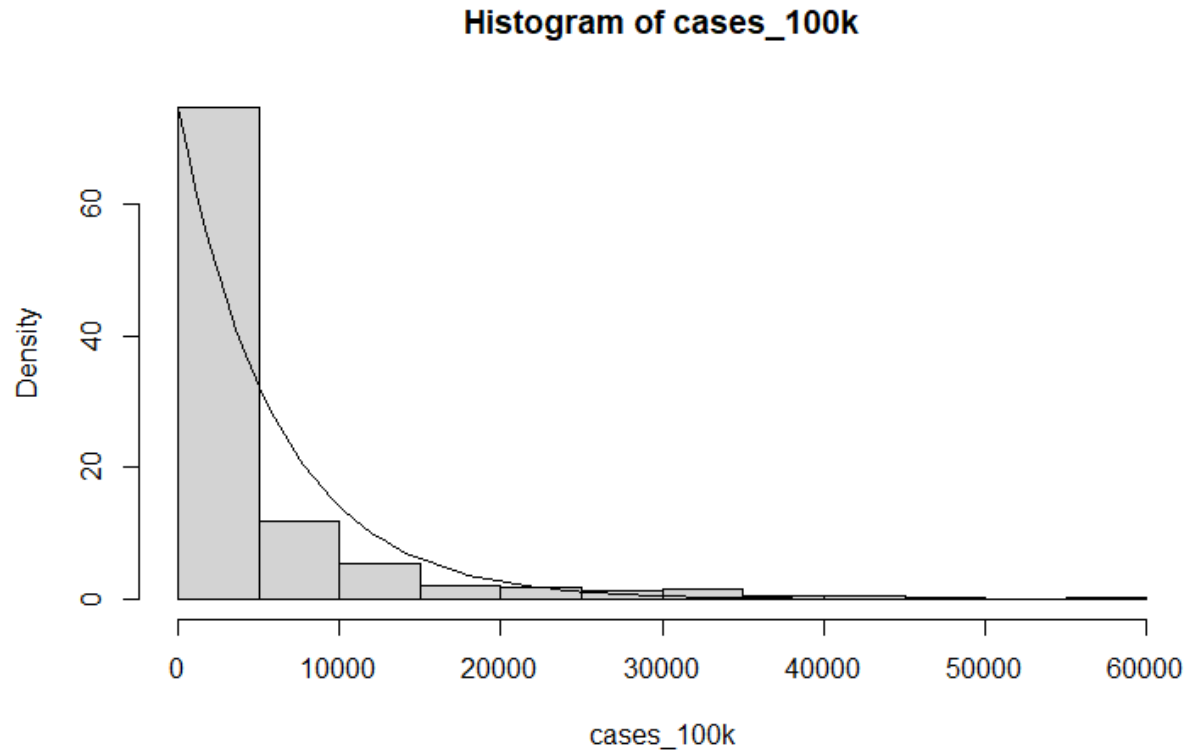


Figure 22 Probability Density for Column Cases_100k

- Inferences:
 - From summary of column, we can see that mean is greater than median which shows positively skewed.

2.4.1.2 Relative Probability Distribution for COVID-19 Dataset & Food Production Dataset

1. **Dataset:** population_by_country_2020 & country_food_data
2. **Column:** meat_per_person
3. **Probability Density function:**

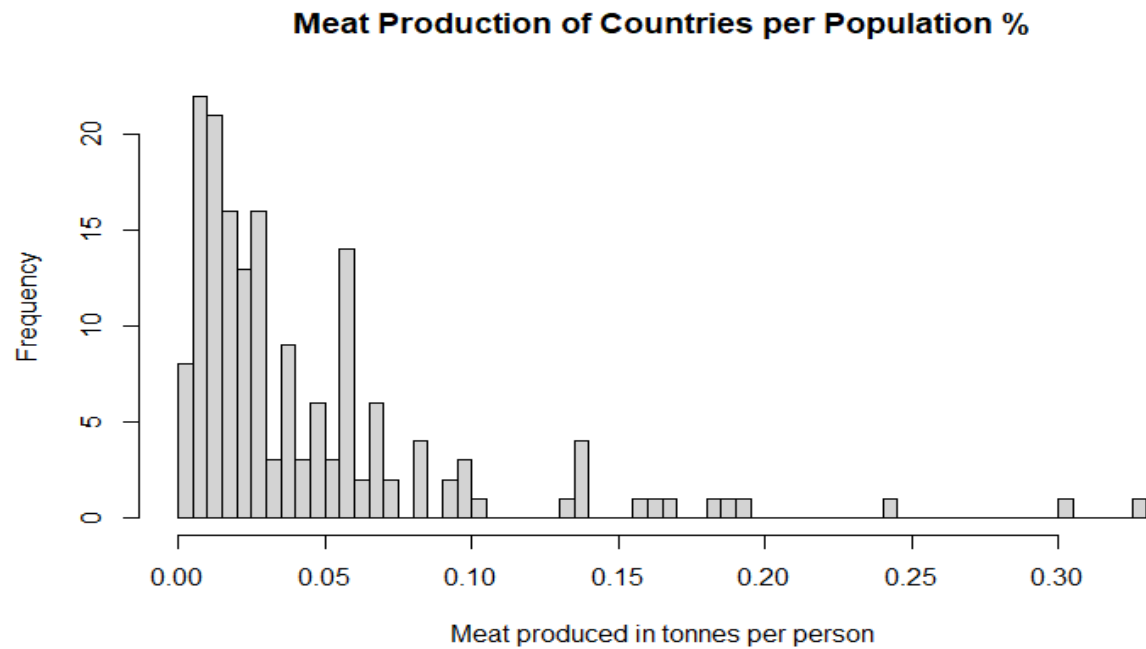


Figure 23 Histogram for Meat production per Population percentage

- Inferences:
 - The number of Item == “Meat, Total” shows positively skewed, geometric distribution.

4. Probability Density for Column meat_per_person:

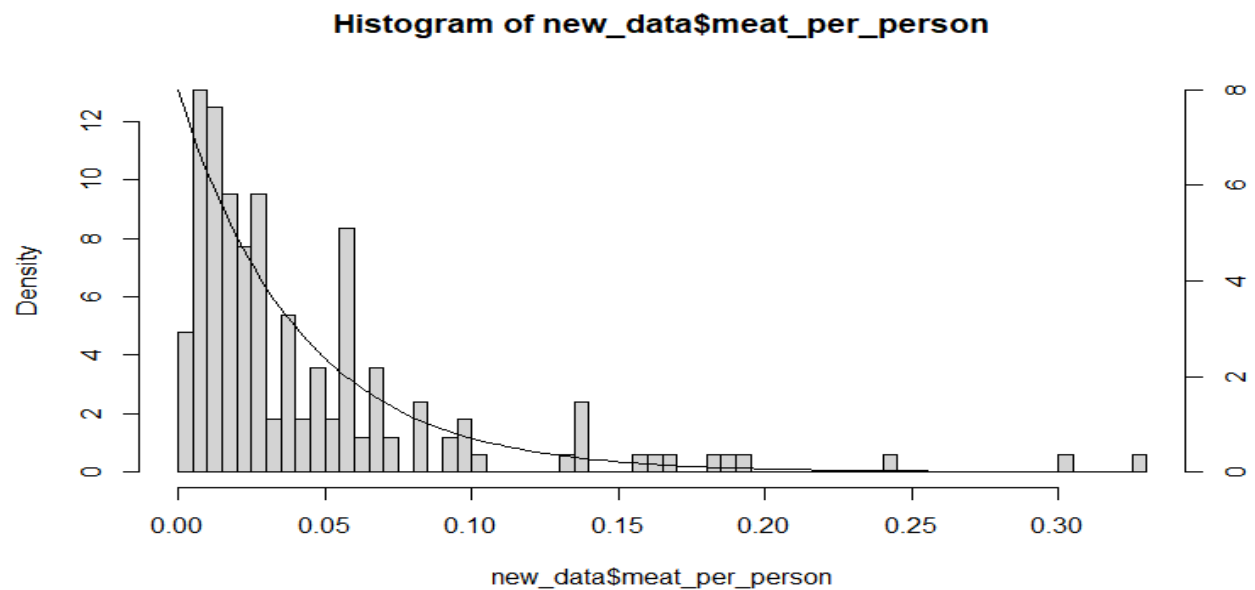


Figure 24 Probability Density for Column Meat production per Population percentage

- Inferences:
 - From summary of column, we can see that mean is greater than median which shows positively skewed.

2.4.1.3 Relative Probability Distribution for Annual Trade Announcements Tally

1. **Dataset:** COVID19_data_cleaned
2. **Column:** Cases_100k
3. **Probability Density function:**

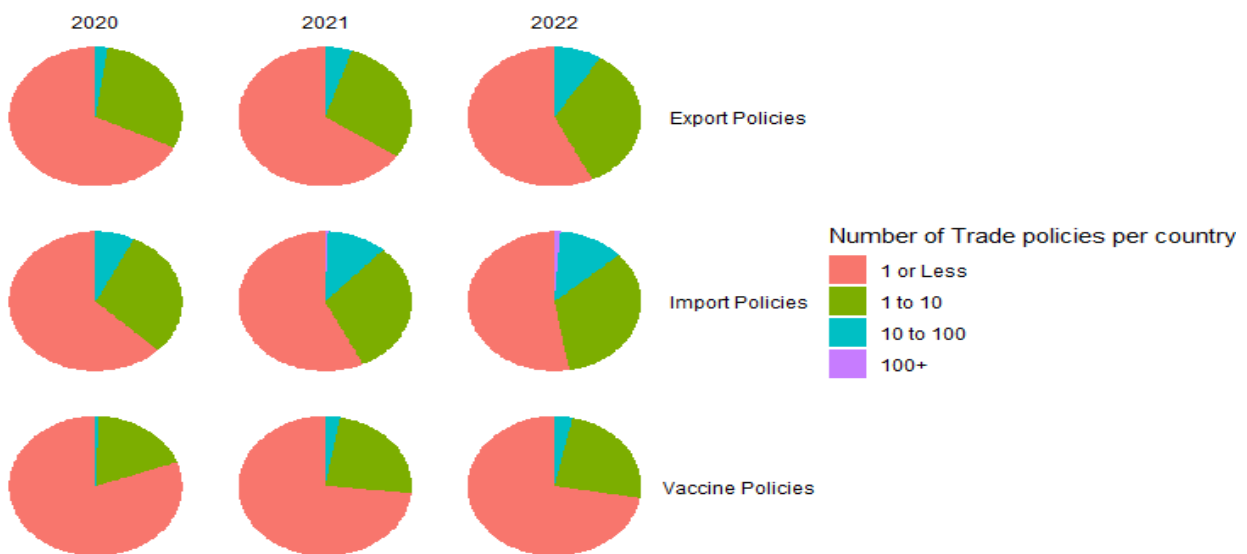


Figure 25 Pie Chart for Number of Trade policies per country

- Inferences:
 - Here, pie chart for year 2020, 2021 and 2022 shows number of trade policies per country first row shows for export policies
 - Here, pie chart for year 2020, 2021 and 2022 shows number of trade policies per country first row shows for import policies
 - Here, pie chart for year 2020, 2021 and 2022 shows number of trade policies per country first row shows for vaccine policies

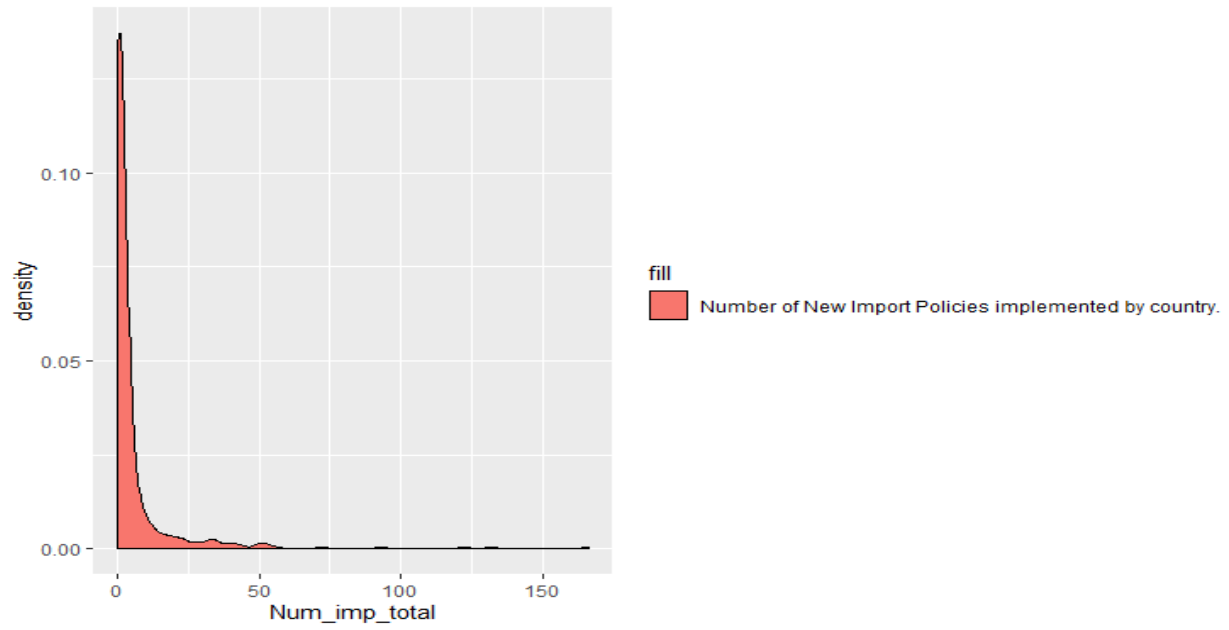


Figure 26 Probability Density for column Number of import total

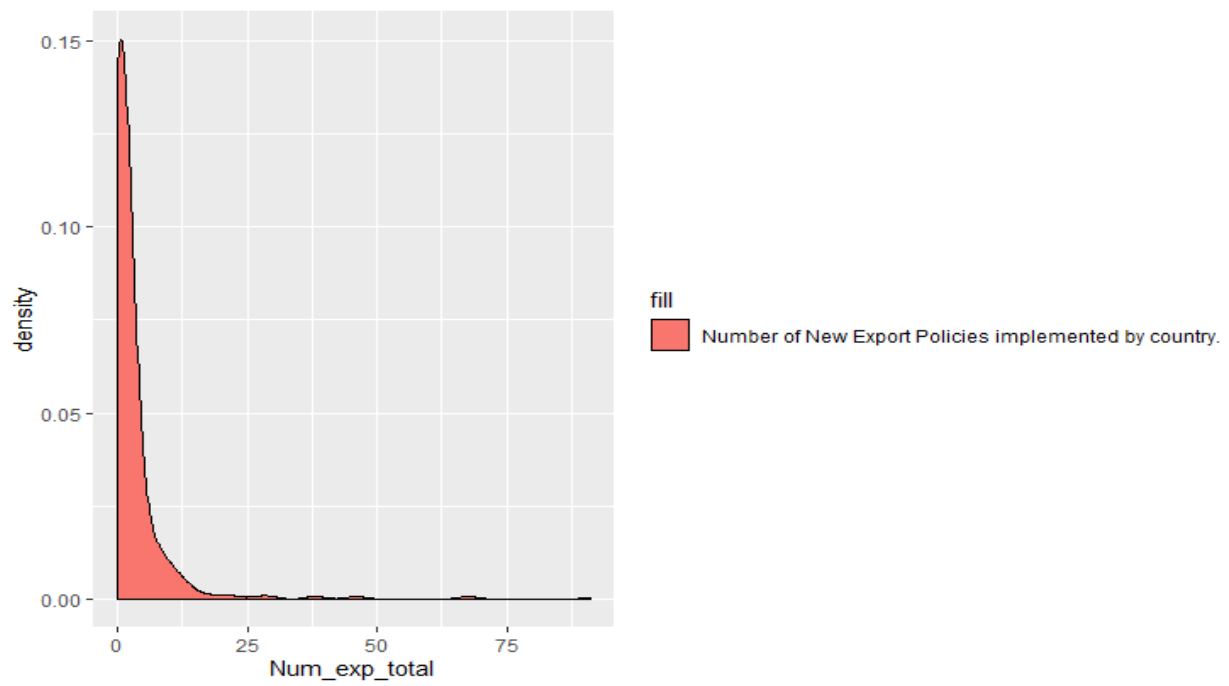


Figure 27 Probability for Column Number of Exports policies implemented

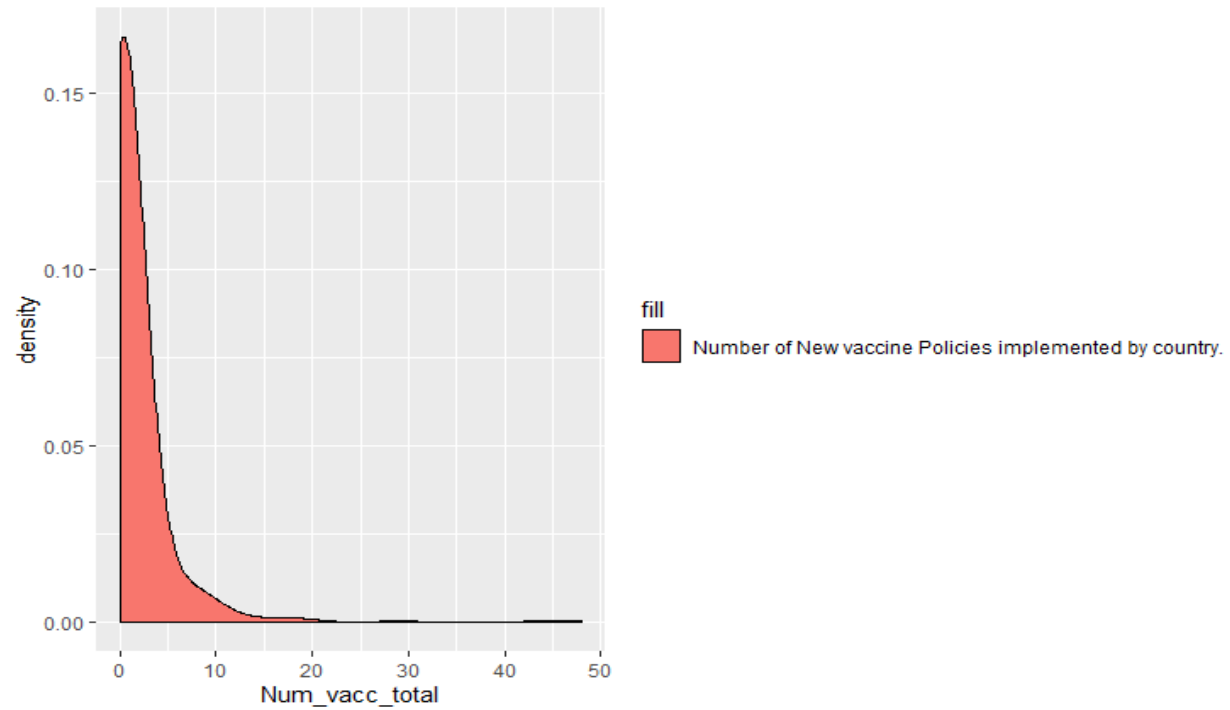


Figure 28 Probability for Column Number of Vaccine policies implemented in total

- Inferences:
 - Column export policies shows positive skewness of data.
 - Column import policies shows positive skewness of data.
 - Column Vaccine policies shows positive skewness of data.

2.4.2. Measure of center and Variation- and Distributions

Two measures of center are mean and median. Spread describes the variation of the data. Two measures of spread are range and standard deviation.

2.4.2.1 Measure of center and Variation for Food Production Dataset & COVID-19

Dataset

1. **Dataset:** Food_production_dataset & CovidAndFood

2. **Column:** Items & New_cases

3. **Summary of Column:**

- Mean: 6775217
- Median: 318010
- Mode: "numeric"
- Range: 24 to 337179926
- Stddev: 27484133
- Variance: 7.553776e+14
- Quantiles:
 - Q1: 65661
 - Q2: 318010
 - Q3: 2486900
- IQR: 2421239

4. **Histograms with different bin sizes**

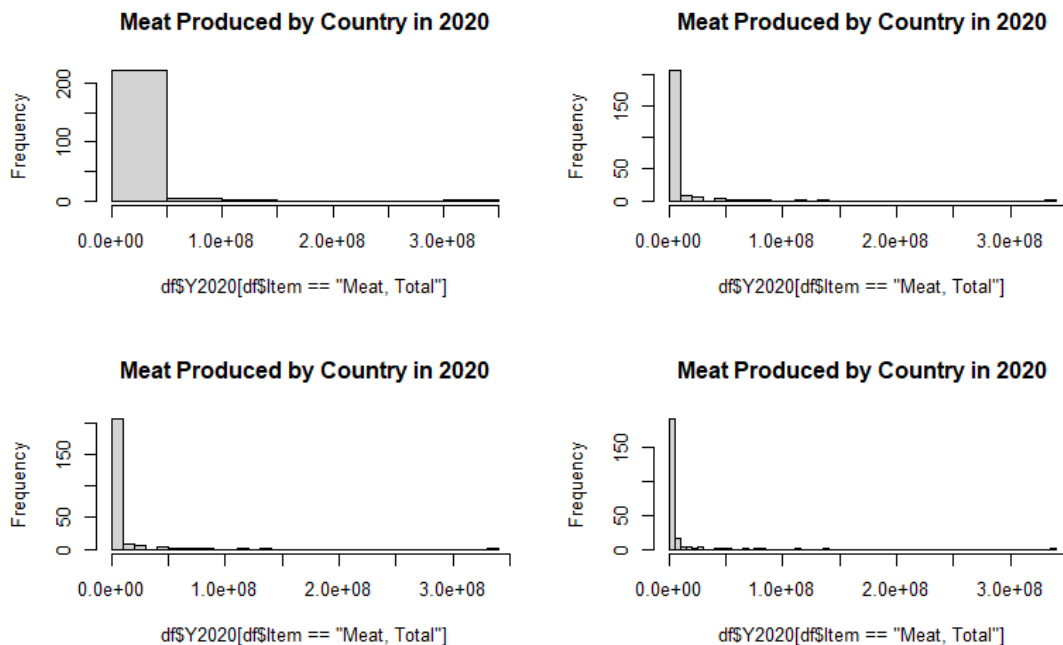


Figure 29 Histogram for column Item == Meat, Total

- Inferences:
 - The number of Item == “Meat, Total” shows positively skewed, geometric distribution.
 - Figure 18 shows different binned sizes for column Item == “Meat, Total” with bins- 5, 25, 40, 50 respectively.
 - From summary of column, we can see that mean is greater than median which shows positively skewed.

5. Boxplot for column Items == “Meat, Total”

Box plot of meat produced by country

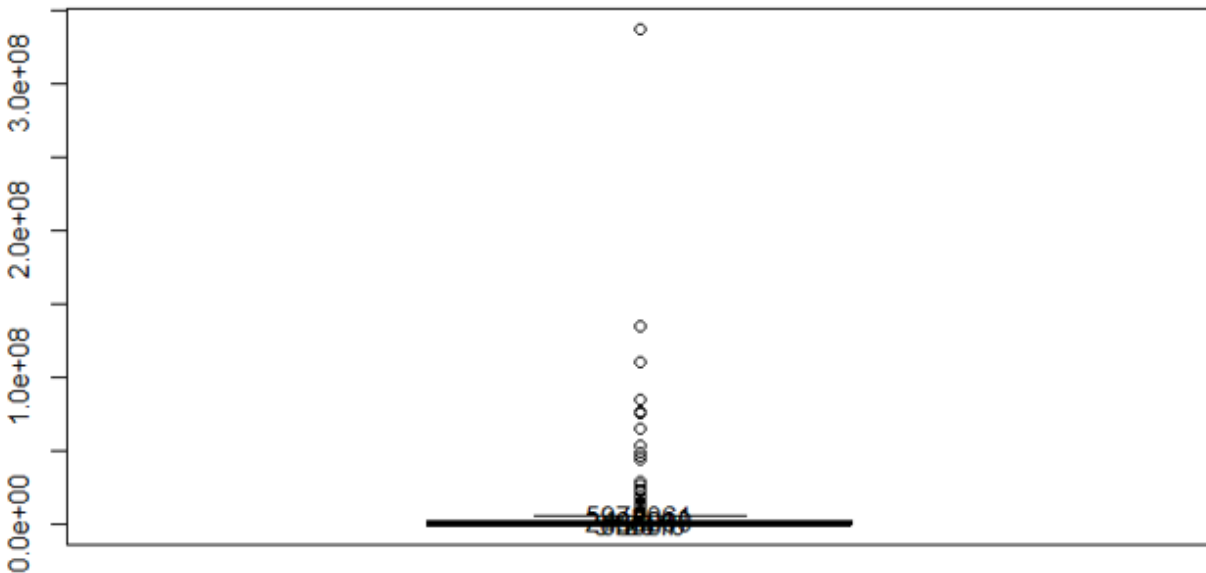


Figure 30 Box plot of meat produced by country

- Inferences:
 - Box plots divide the data into sections that each contain approximately 25% of the data in that set.
 - Here, data is concentrated less than 10K and shows many outliers.

6. KDE plot for 2020 cases

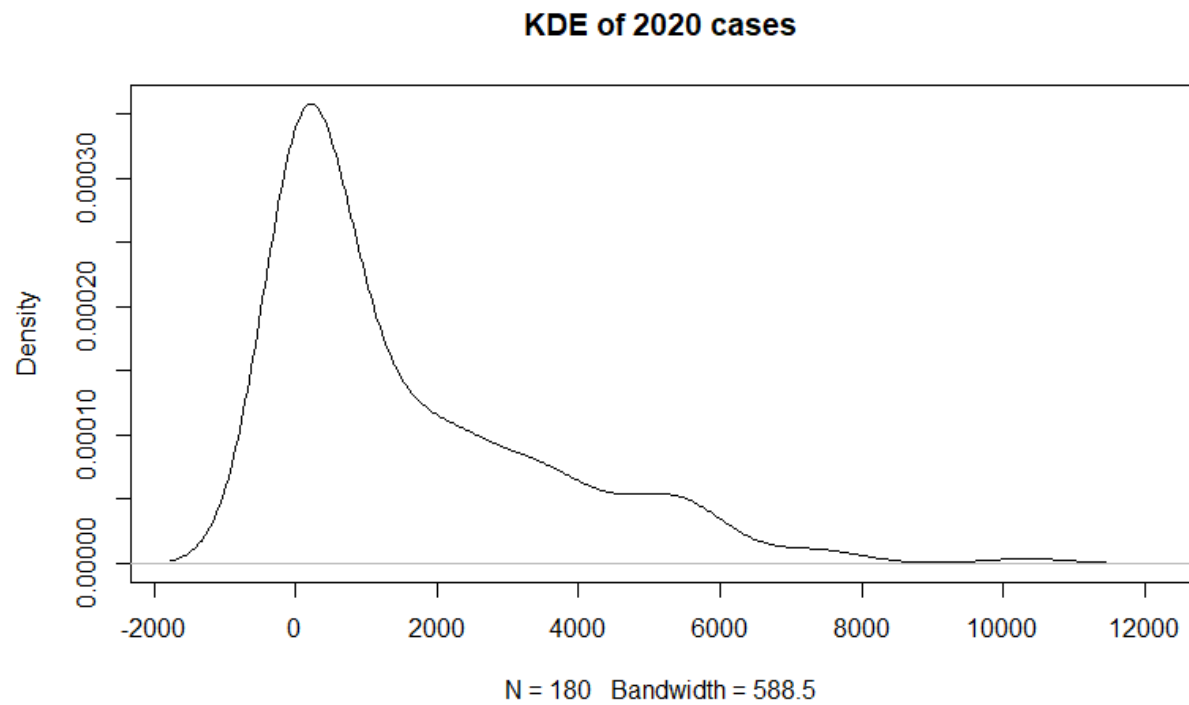


Figure 31 KDE for 2020 cases (Gaussian)

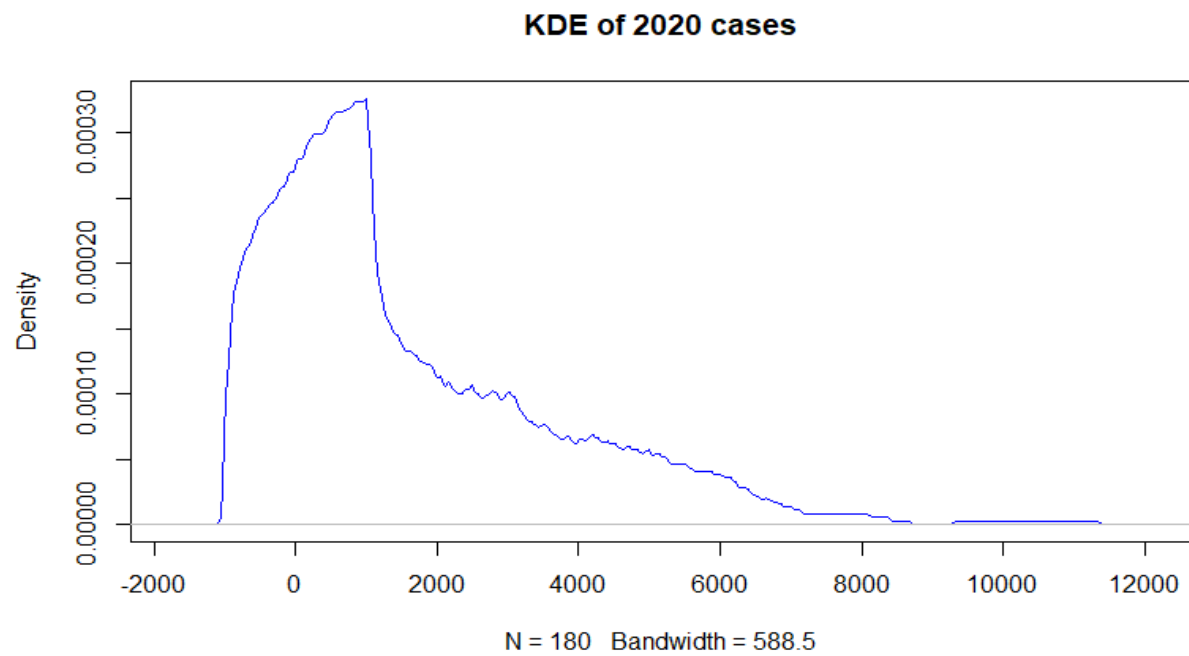


Figure 32 KDE for 2020 cases (Rectangular)

- Inferences:
 - The density plot for 2020 cases, shows similar distribution to the boxplot.
 - Changing the KDF from gaussian to other values allows new ways to look at the data.
 - Gaussian and rectangular Kernel is used in these graphs.
 - Both KDE shows the skewness of data on right.

7. Q-Q plot for Covid cases in 2020

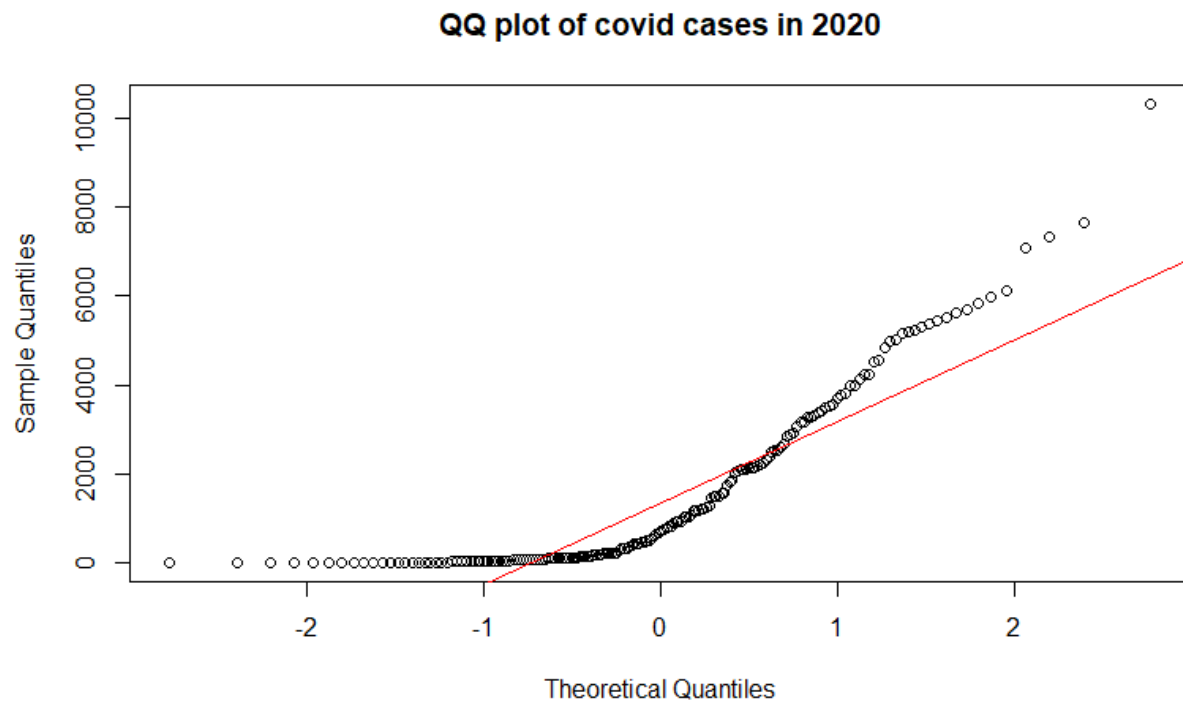


Figure 33 Q-Q plot for COVID cases in 2020

- Inferences:
 - The Q-Q plot for Covid cases in 2020, show geometric nature of the variable distribution.
 - The Q-Q plot for Covid cases in 2020, show linear relationship, also, positive correlation.

2.4.2.2 Measure of center and Variation for AnnualTradeAnnouncementsTally

1. **Dataset:** AnnualTradeAnnouncementsTally
2. **Column:** new Trade Policies announced by country and year (2020-2022)
3. **Summary of Column:**
 - Mean: 6.73125
 - Median: 0
 - Mode: 0
 - Range: 0 to 283
 - Variance: 568.7568
 - Stddev: 23.84862
 - Quantiles:
 - Q1: 0
 - Q2: 0
 - Q3: 2
 - IQR: 2
4. **Histograms with different bin sizes:**

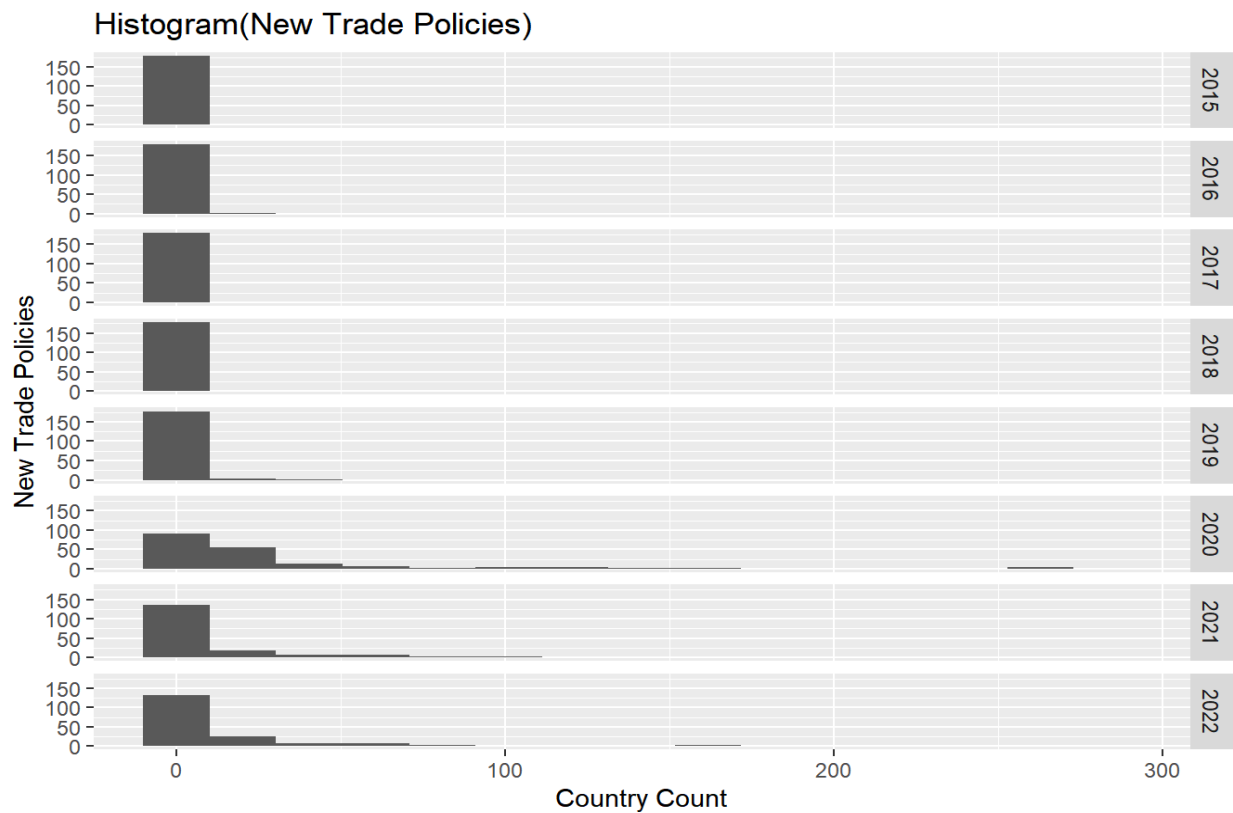


Figure 34 Histogram New Trade Policies Yearly

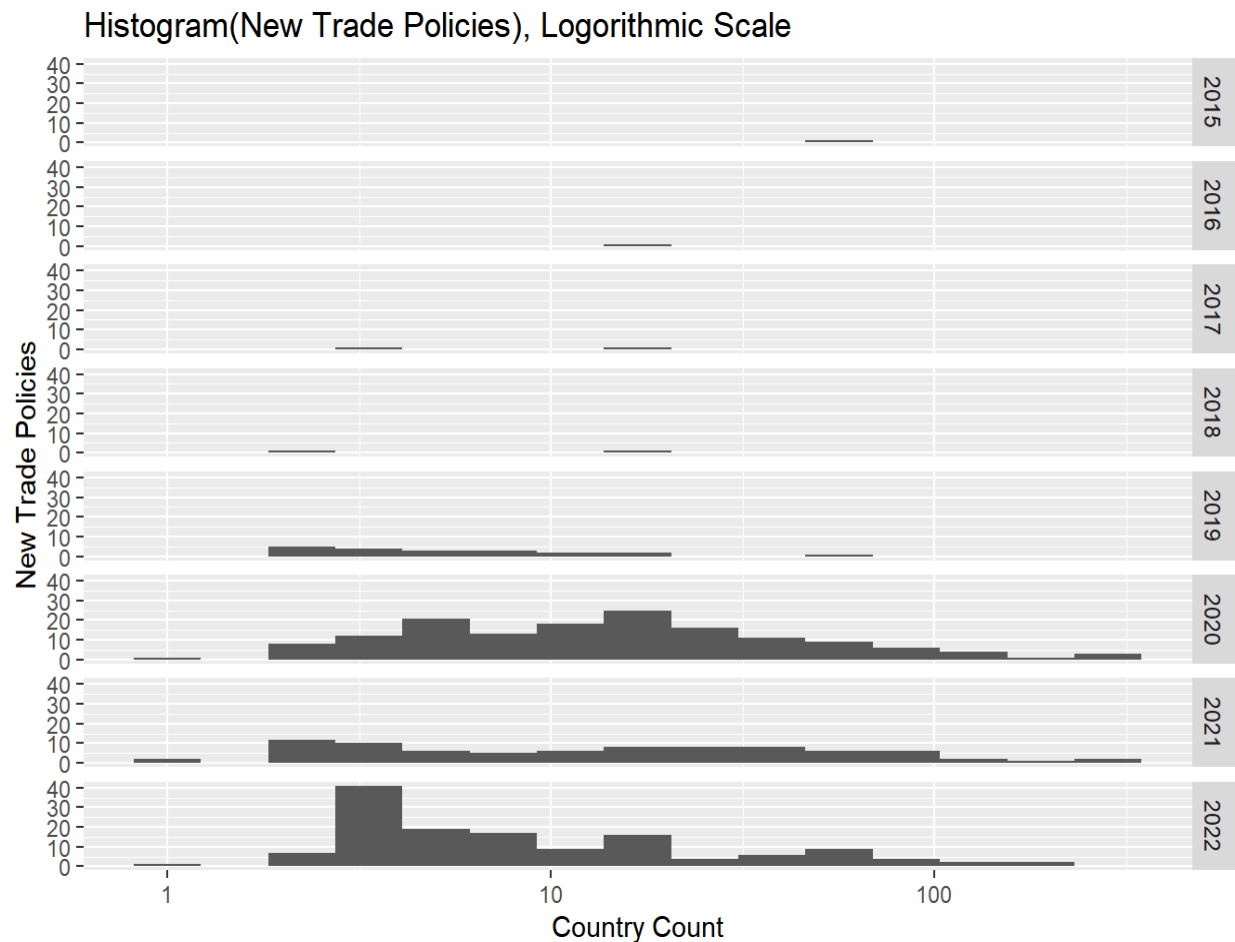


Figure 35 Histogram New Trade Policies Yearly (Logarithmic Scale)

- Inferences:
 - The number of new trade policies issued per country-year displays a positively skewed, geometric distribution.
 - Most countries issued no new recorded trade policies from the years of 2021-2022, and the number of trade policies issued by countries which did issue new policies is distributed geometrically. Nevertheless, the number of trade policies issued by each country increased dramatically in 2020, after the onset of COVID-19. The geometric nature of this distribution persists from year to year, even as the mean, median, and IQRs.

5. Violin plots for New trade Policies:

Violin Plots of New Trade Policies by Year

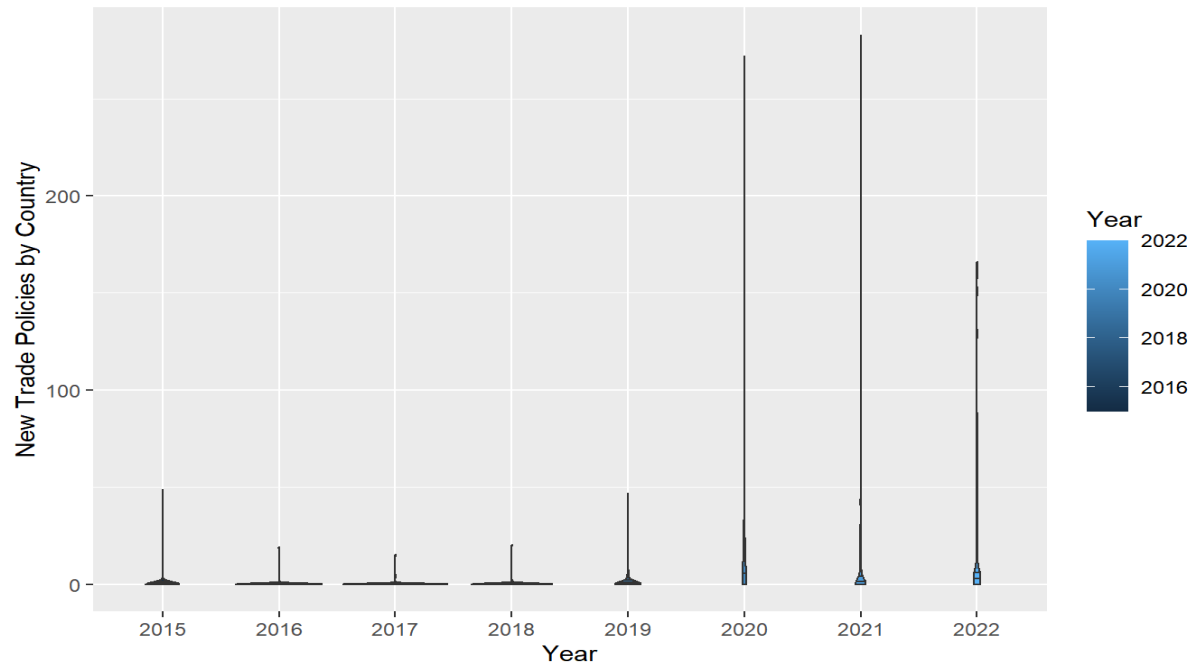


Figure 36 Violin plot for New Trade policies by year

Violin Plots of New Trade Policies by Year, Logarithmic Scale

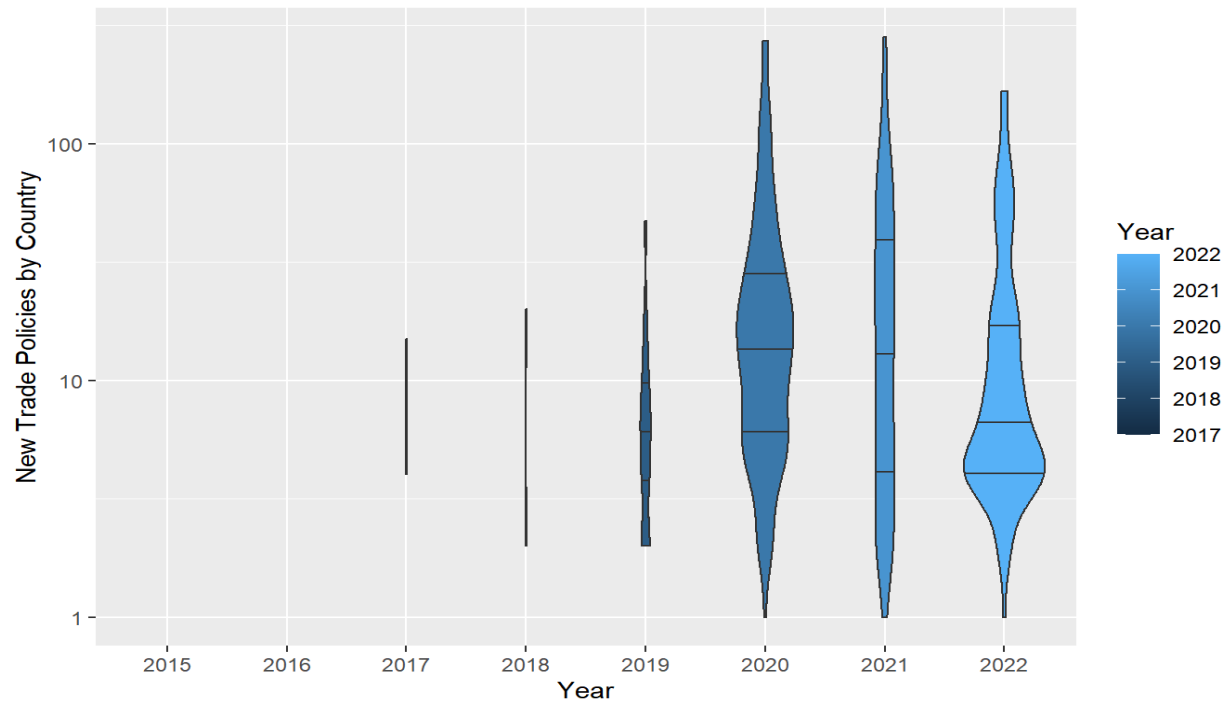


Figure 37 Violin Plots of New Trade Policies by year, Logarithmic Scale

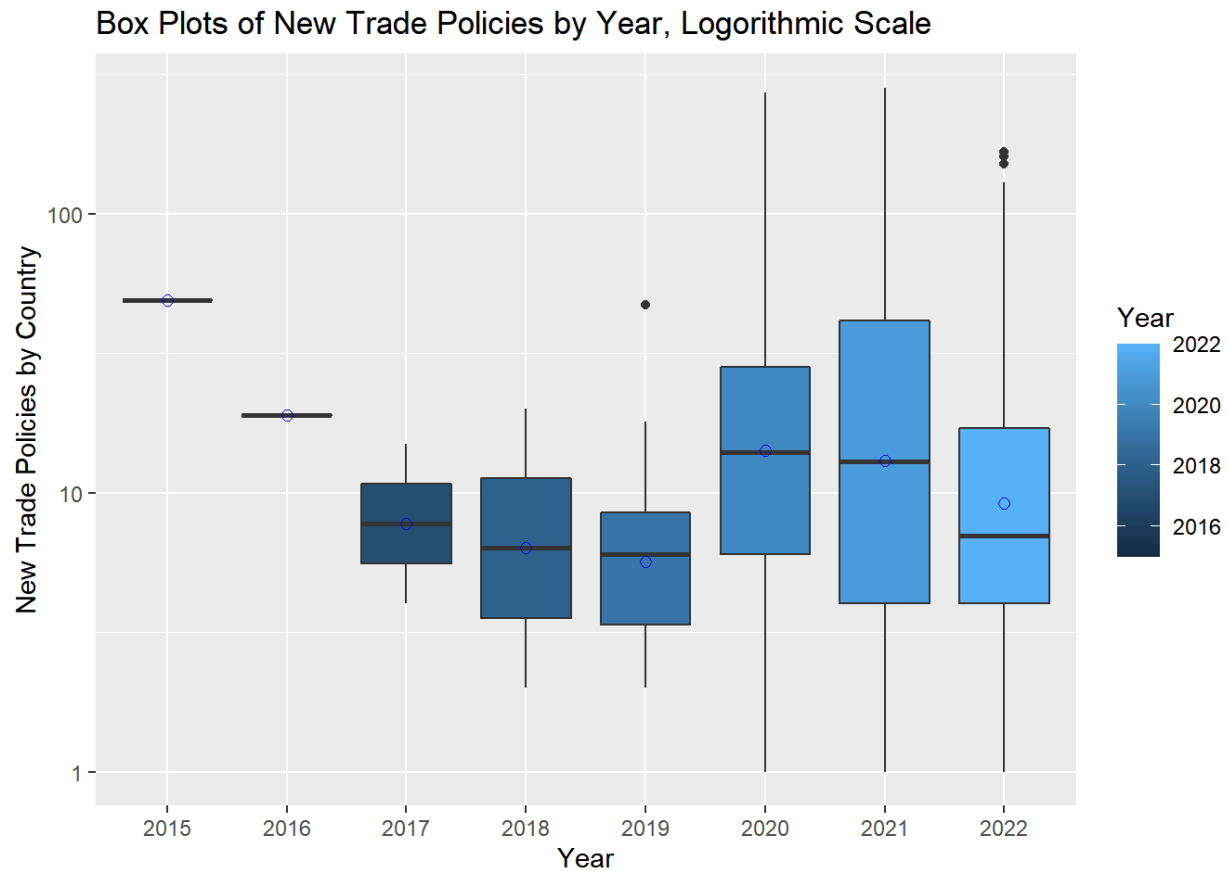


Figure 38 Boxplot of New Trade Policies by Year, Logarithmic scale

- Inferences:
 - Plots show a distinct, sharp increase in the number of trade policies enacted by each country during the years of 2020-2022.
 - The violin plots show a more distinct trend, where the year 2020 saw the greatest mean # of new trade policies per country, 2021 saw a mixed distribution, and 2022 saw a lower mean # of new trade policies (though still significantly more than in the years prior to 2020).
 - Blue circles mark the mean new trade policies per country, while the lines on the violin plot mark the 1st, 2nd, and 3rd quartiles, when read from bottom up.

6. KDE plot for New Trade policies:

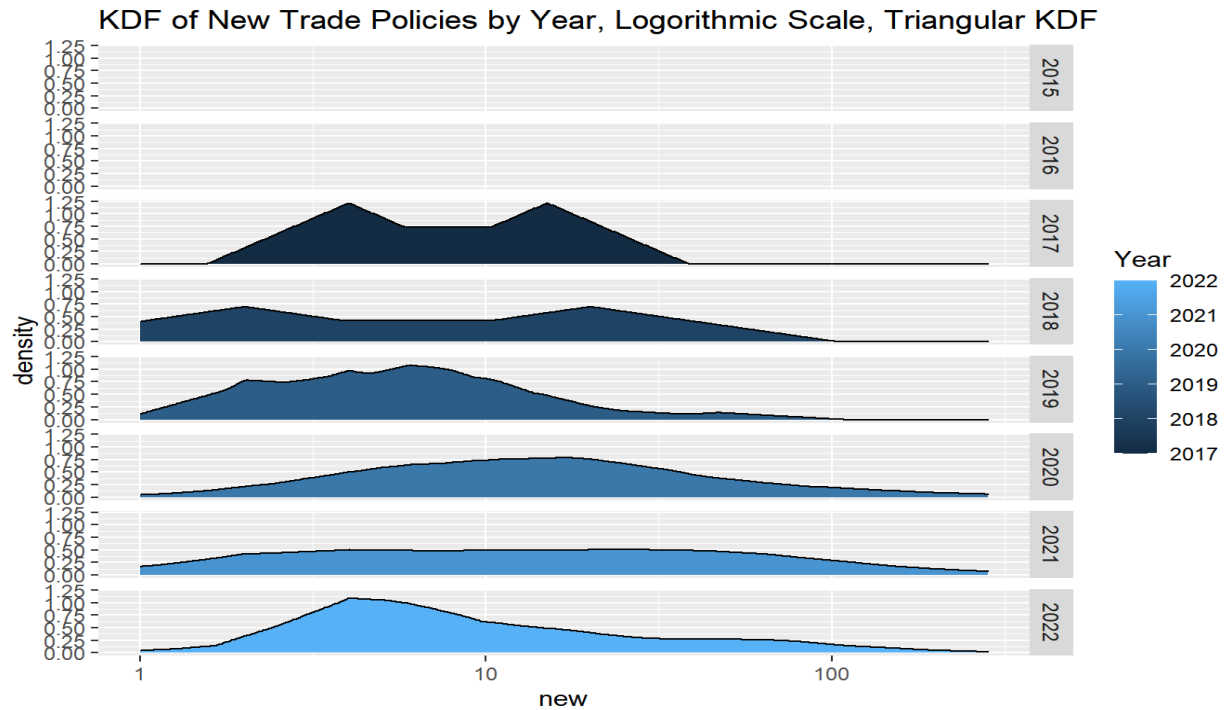


Figure 39 KDF of New Trade Policies by year, Log scale, Triangular KDF

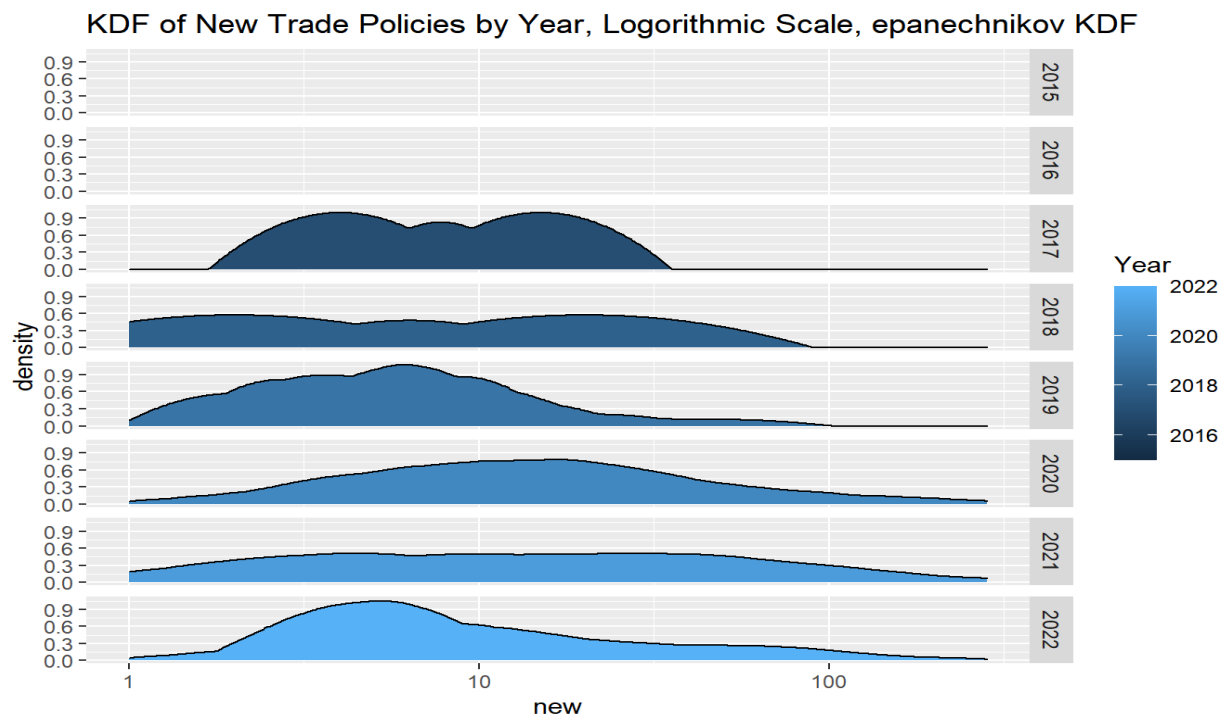


Figure 40 KDF of New Trade Policies by year, Log scale, epanechnikov KDF

- Inferences:
 - The density plot of trade policies per year shows a similar distribution to the violin plots.
 - Changing the KDF from a gaussian to other values allows new ways to look at the data. Triangular and epanechnikov KDFs were used for these graphs.

7. Q-Q plot for New Trade policies:

Q-Q Plot of New Trade Policies Per Country, grouped by Year.

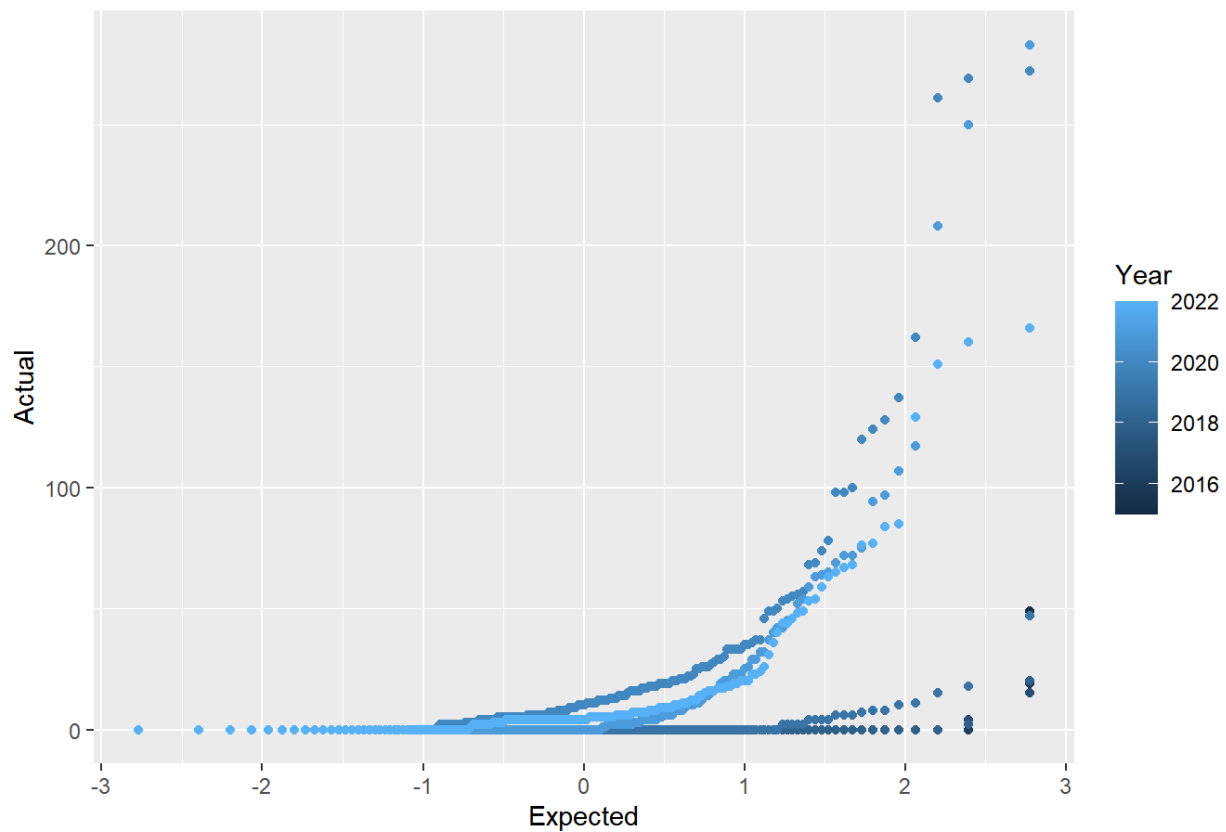


Figure 41 Q-Q Plot for New Trade Policies per Country, grouped by year

- Inferences:
 - The QQ Plot of the new trade policies issued per country per year clearly shows the geometric nature of the variable distribution. Rather than a linear relation between the expected and actual variables, there is a strong non-linear relation for all years.
 - However, the years of 2020, 2021, and 2022 show a high degree of positive skew at the top end. They reflect a similar geometric distribution as the previous years, but on a larger scale.

3. Results:

Hypothesis testing is the process used to evaluate the strength of evidence from the sample and provides a framework for making determinations related to the population, i.e., it provides a method for understanding how reliably one can extrapolate observed findings in a sample under study to the larger population from which the sample was drawn. The investigator formulates a specific hypothesis, evaluates data from the sample, and uses these data to decide whether they support the specific hypothesis.

3.1 Food Security Environment Scores 2017 & 2021

Data is from numerous different sources. The first one is one that already have been used, the CovidAndFood data, but here using the covid part of the data and looking at new cases in each country. The second data is the "Global Food Security Index 2022" by Economist Impact. The scores of each country are normalized from 0-100, where 100 are the best possible conditions." The Global Food Security Index (GFSI) considers food affordability, availability, quality and safety, and sustainability and adaptation across 113 countries. " This description is from the 2022 report. Separation of the countries by their income group according to that same paper, 19 countries go into the "Low Income Group" and 38 countries go into the "High Income Group." Here is how they land on the spread of Food Security from the years 2017 and 2021, with low income being in red, high income being in blue, and the rest of the population of countries from this survey in grey.

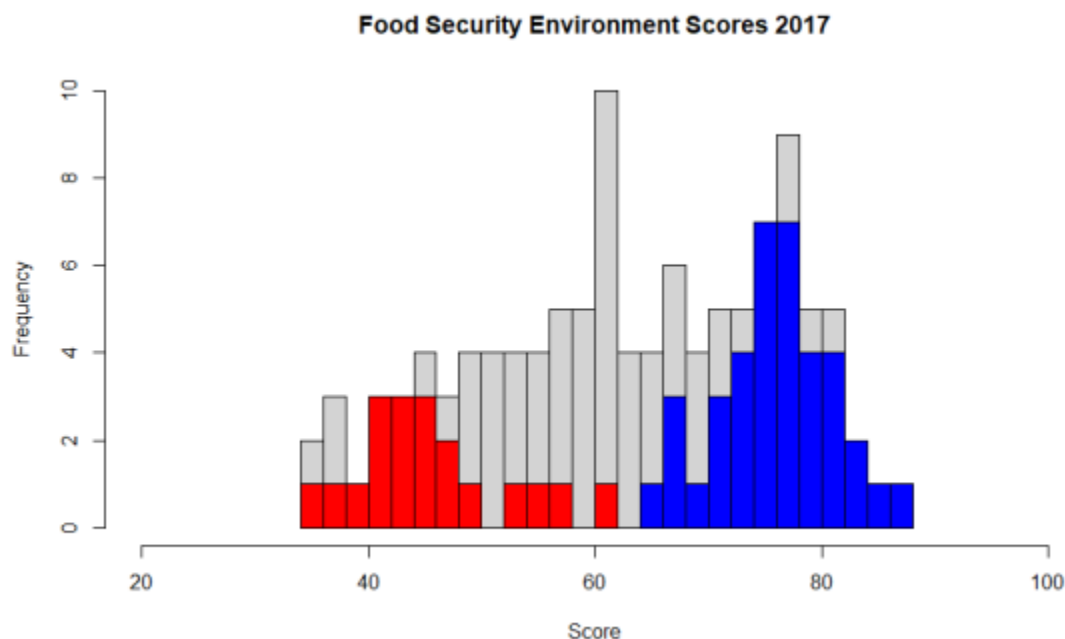


Figure 42 Food Security Environment Scores 2017

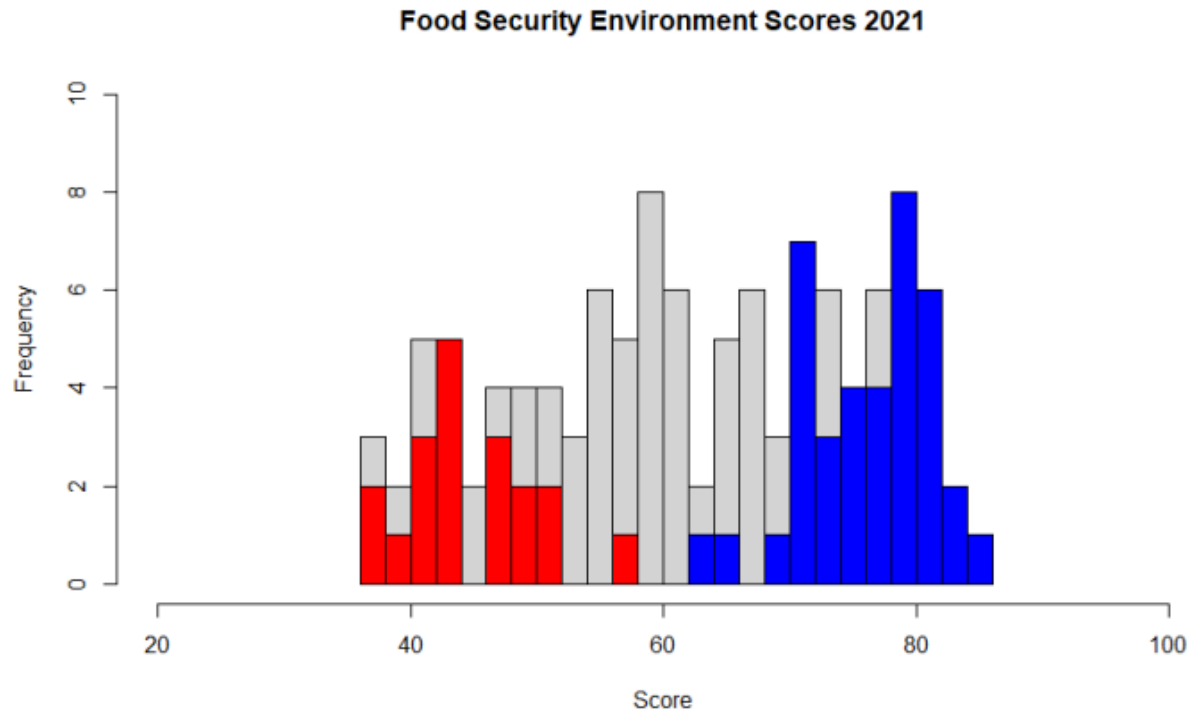


Figure 43 Food Security Environment Scores 2021

We can see that the low-income countries are on the lower end and high-income countries opposite. Is the average food security of those groups significantly different from the mean? For that we can use some hypothesis testing. Here we are doing hypothesis testing on two years, one from before COVID-19 and one from after, specifically 2017 and 2021.

```

> summary(study)
      x2017      x2021
Min.   :34.30  Min.   :36.40
1st Qu.:52.60  1st Qu.:51.80
Median :63.20  Median :61.70
Mean   :62.74  Mean   :62.13
3rd Qu.:74.50  3rd Qu.:73.90
Max.   :86.50  Max.   :84.70
> summary(hic_foodsecurity)
      x2017      x2021
Min.   :65.70  Min.   :62.40
1st Qu.:73.12  1st Qu.:72.28
Median :76.35  Median :76.75
Mean   :76.07  Mean   :75.90
3rd Qu.:79.38  3rd Qu.:79.33
Max.   :86.50  Max.   :84.70
> summary(low_foodsecurity)
      x2017      x2021
Min.   :34.30  Min.   :36.40
1st Qu.:41.45  1st Qu.:41.70
Median :45.10  Median :43.90
Mean   :45.56  Mean   :44.77
3rd Qu.:48.40  3rd Qu.:47.90
Max.   :60.20  Max.   :56.60
> sd(study$x2017)
[1] 13.41319
> sd(study$x2021)
[1] 13.21168
> sd(low_foodsecurity$x2017)
[1] 6.859722
> sd(low_foodsecurity$x2021)
[1] 5.194447

```

Figure 44 Summaries of Year 2017 & 2021

From the summaries, we can see that the skew of all countries in total shifts from slightly right to slightly left going from 2017 to 2021 when comparing the median and mean values. The high-income countries, or hic food security in the picture, doesn't change as much in this timeframe and the low-income countries change a lot more comparatively. Is this change significant?

For that we are doing 4 hypothesis tests—two on each of the groups with $\alpha = 0.01$.

Doing Z tests for the high-income group and T tests for the low-income group due to the sample size of them (high income has 38 countries while low income only has 19).

- H_0 : The null hypothesis for all these tests is going to be that the mean food security score of the selected group is equal to the mean food security score of the population of given countries.
- H_a : The alternative hypothesis is that the higher income countries have a higher mean than the population mean, and the low-income countries are lower.

```

> summary(df2$Cases_100k[df2$Year=='2020'])
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.   NA's
    0.0   85.1   683.7  1601.5  2560.7 10332.0    57
> summary(hic_Covid$Cases_100k[hic_Covid$Year=='2020'])
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.   NA's
   35.5  1288.9  3260.8  2865.4  4128.1  5616.8     1
> summary(lic_Covid$Cases_100k[lic_Covid$Year=='2020'])
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.   NA's
   6.879  32.307  50.886  53.430  67.227  111.304     1
> sd(df2$Cases_100k[df2$Year=='2020'], na.rm =TRUE)
[1] 1989.288
> sd(lic_Covid$Cases_100k[lic_Covid$Year=='2020'], na.rm =TRUE)
[1] 33.57218

```

Figure 45 Summary of Countries cases per 100K people from 2020

This is data on those same groups of countries for covid cases per 100 thousand people from 2020. Just by looking at the means we can see that they are very different from the population mean, but we will still run tests on them. The number of countries changed a little due to having a different dataset, so the high-income country only has 34 instead of 38.

- Z-test :

$$z = \frac{(\bar{X} - \mu)}{\frac{\sigma_{pop}}{\sqrt{n}}}$$

$$Z_{2017} = \frac{76.07 - 62.74}{\frac{13.41}{\sqrt{38}}}$$

$$Z_{2017} = \frac{13.33}{2.17}$$

$$Z_{2017} = 6.127$$

$$Z_{2021} = \frac{75.90 - 62.13}{\frac{13.21}{\sqrt{38}}}$$

$$Z_{2021} = \frac{13.77}{2.14}$$

$$Z_{2021} = 6.425$$

$$Z_{CH} = \frac{2864 - 1601}{\frac{1989}{\sqrt{34}}}$$

$$Z_{CH} = \frac{1263}{341.11}$$

$$Z_{CH} = 3.702$$

○ t-test:

$$T = \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}}$$

$$T_{2017} = \frac{45.56 - 62.74}{\frac{6.859}{\sqrt{19}}}$$

$$T_{2017} = -\frac{17.18}{1.57}$$

$$T_{2017} = -10.9179$$

$$T_{2021} = \frac{44.77 - 62.13}{\frac{5.194}{\sqrt{19}}}$$

$$T_{2021} = -\frac{17.36}{1.57}$$

$$T_{2021} = -14.5688$$

$$T_{CL} = \frac{53 - 1601}{\frac{33.5}{\sqrt{19}}}$$

$$T_{CL} = -\frac{1548}{7.685}$$

$$T_{CL} = -201.42$$

For the high-income group, we use their means 76.07 and 75.90 against their respective population means of 62.74 and 62.13. The standard deviations are 13.41 for the year 2017 and 13.21 for 2021. With those values we get z scores of 6.127 for 2017 and 6.425 for 2021. Both of those are well above the threshold to be considered significantly different, so we will reject then null hypothesis! For the low-income group, the means are 45.56 for 2017 and 44.77 for 2021. Because of the smaller sample size, the use of sample standard deviation is needed, which is 6.859 for 2017 and 5.194 for 2021. With these and the population means we get T values of -10.9179 for 2017 and -14.5688 for 2021. We have 18 degrees of freedom, so the critical value for one tailed T tests with that many degrees of freedom and $\alpha = 0.01$ is -2.552. We can see that both years they are significantly different, so we can reject the null hypothesis. Looking at the values from the tests, we can also see that the low-income countries changed a lot more comparatively to the high-income countries. In other words, their food security was affected more during COVID. For the covid cases, we do the same steps of Z test for high income and T test for low. We end up with scores of 3.70 for high income countries and -201.42 for low-income countries. These are both statistically significant. While this shows that there were less deaths on average in lower income countries, We believe that isn't because of their income but rather that covid affected more highly dense populations, which higher income countries usually are, and low-income countries aren't.

3.2 Trade Policies before and after COVID-19

International trade plunged in 2020 but recovered sharply in 2021. While total trade flows are now comfortably above pre-pandemic levels, trade impacts across specific goods, services and trade partners are highly diverse, creating pressures on specific sectors and supply chains. The changes in the trade structure caused by the COVID-19 pandemic in a single year was of a similar magnitude to changes otherwise typically seen over 4-5 years. Substantial imbalances across trade partners and products remained at the end of 2021, and not all the accumulated losses from the earlier steep declines were recuperated. The heterogeneity of trade impacts and changes in trade flows across products, sources and destinations signifies high uncertainty and adjustment costs, and implies additional incentives for consumers, firms, and governments to adopt new — or to intensify existing — risk mitigation strategies.

- Ha: The number of trade policies issued globally each year was significantly altered by the outbreak of Covid in the beginning of 2020.
- Ho: The number of trade policies issued globally each year was not significantly altered by the outbreak of Covid in the beginning of 2020

To determine whether the observed counts are significantly different from the counts that we would expect if there was no association between the two variables. We have the observed counts, so we now need to compute the expected counts in the case the variables were independent.

These expected frequencies are computed for each subgroup one by one with the following formula:

$$\text{Exp.Frequencies} = \frac{\text{total \# of obs. for row} \cdot \text{total \# of obs. for the column}}{\text{total number of observations}}$$

To calculate trade frequency, we will calculate the sum of columns 'Num_exp_total' and 'Num_imp_total'. We have the observed and expected frequencies. We now need to compare these frequencies to determine if they differ significantly. The difference between the observed and expected frequencies, referred as the test statistic (or t-stat) and denoted χ^2 , is computed as follows:

$$\chi^2 = \sum_{i,j} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

where O represents the observed frequencies and E the expected frequencies. We use the square of the differences between the observed and expected frequencies to make sure that negative differences are not compensated by positive differences.

The critical value can be found in the statistical table of the Chi-square distribution and depends on the significance level, denoted α , and the degrees of freedom, denoted df. The significance level is usually set equal to 5%. The degrees of freedom for a Chi-square test of independence are found as follow:

$$df = (\text{number of rows} - 1) \cdot (\text{number of columns} - 1)$$

We calculate chisq test using R language. We get parameter numbers as 7 , sum of trade frequency is N as 4363 chi- statistics is 4692.55 and p-value as 0.0000. By the Chi Squared test, the distribution of trade policies is clearly not symmetric, with $P < .00001$. Thus, the number of trade policies issued varies significantly. Hence, we reject null hypothesis.

3.3 Trade Policies announced globally

A t-test is a statistical test that is used to compare the means of two groups. It is often used in hypothesis testing to determine whether a process or treatment actually has an effect on the population of interest, or whether two groups are different from one another.

Defining Hypothesis assumptions as follow:

- Ho: The annual number of trade policies announced globally was not significantly different before / after the outbreak of Covid.
- Ha: The annual number of trade policies announced globally was significantly different before / after the outbreak of Covid.

We calculate t-test using R language. We are running a 2-factor t test to estimate how COVID modulates the trade policy distribution. We get "T-test(2.005087,N=9) = -4.98, p = 0.0379"

To conclude The number of trade policies issued globally by year was significantly affected by the outbreak of Covid by a t-test, $p < .05$ Hence We reject the Null Hypothesis.

Following graph shows change in trade policies pre/post-covid.

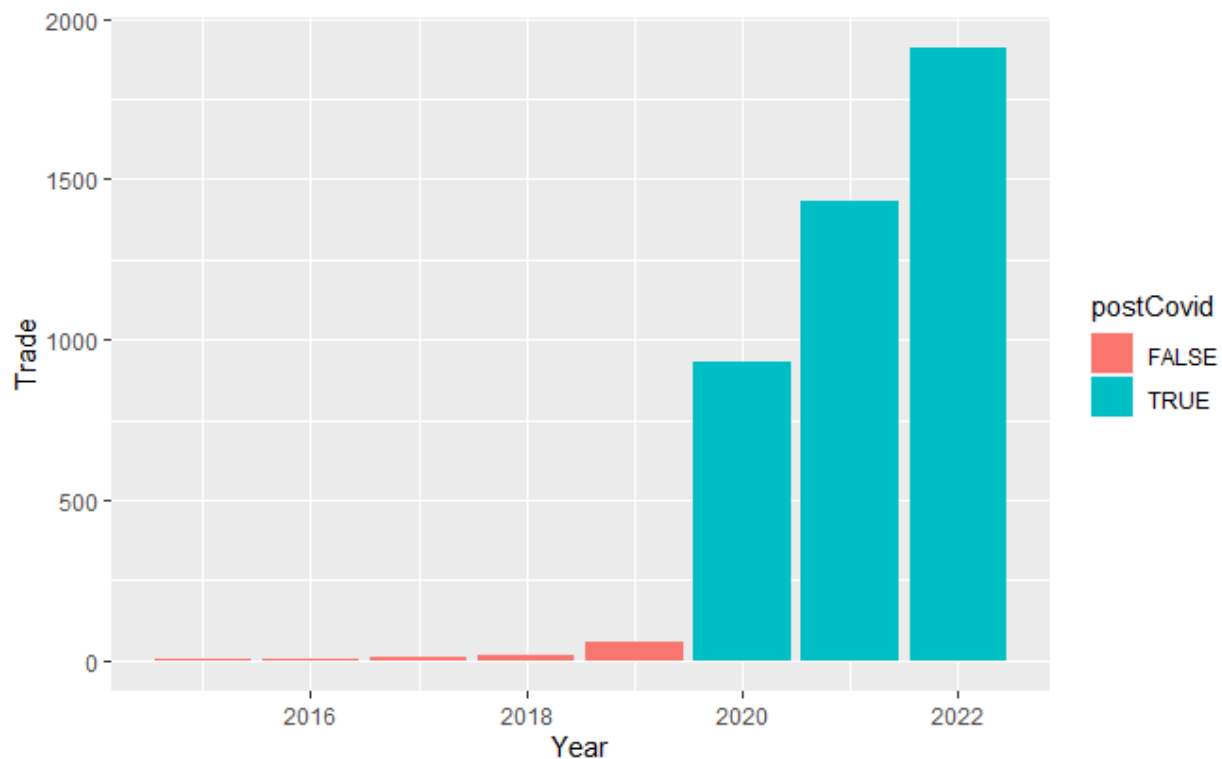


Figure 46 Change in Trade Policies Pre/Post Covid

4. Conclusion:

International trade in 2021 has recovered sharply from the slump in 2020. Despite impressive growth rates of world trade flows, the accumulated losses were not yet recuperated by the end of 2021, but the gap can be expected to close in the first quarter of 2022. While total trade flows are now comfortably above pre-pandemic levels, trade impacts across specific goods, services and trade partners are highly diverse and have been creating pressures on specific sectors and supply chains. Substantial imbalances across trade partners and products remained at the end of 2021, most notably an increased merchandise trade surplus in Asia and a widening merchandise trade deficit in the United States as well as in Africa.

Along with the increase in trade flows during this period, the food volatility also increased. We saw in the hypothesis testing that while more developed countries were not affected by the pandemic when it came to food security, the more developed countries were and were affected by a large margin. This change is due to a combination of relying on more of those outside trades for food, disruption of the production within their own borders, and the change in price volatility of food. Food was impacted to a much greater extent even though not as many people in those developing countries caught or died from the pandemic. If the international trades weren't going to Africa, where most of the developing nations tested are, this situation might have been much worse. While there was still a trade deficit in Africa, it is understandable because their production was affected more than most.

While it is still unknown which of the structural changes seen in 2020 and 2021 will only be short-lived, some seem to suggest longer-term shifts or seem likely to result in long-term adjustments. The pronounced shift of consumer expenditures towards 'home nesting' goods and away from certain services that require person-to-person interaction is unlikely to persist. On the other hand, the big digitalization push that materialized both in the work sphere and personal lives can be expected to have lasting impacts on the composition of demand for products and services and the way those are traded internationally. The unprecedented heterogeneity of changes in trade flows across products, sources and destinations signifies high uncertainty and adjustment costs, and implies additional incentives for consumers, firms, and governments to adopt new—or to intensify existing—risk mitigation strategies. Some firms may want to rethink the resilience and reliability of their supply chains and may decide to try to shorten distances travelled from factories to consumers or internalize larger segments of their value chains within their own corporate structures (e.g. an affiliate supplying a component rather than an external firm). This might contribute to resilience of some supply chains, but it might also have negative impacts on productivity, it may not necessarily boost systemic resilience and stability of the global economy.

Volatility, which is higher in low-income countries, is expected to persist in the medium term due to multiple global and domestic factors. Structural factors contributing to the volatility include rising populations and changing diets, increasingly intertwined relations between food and energy prices. While a troubled global economy could dampen demand and push food prices down,

the effect on developing countries would be mixed—hurting food exporting countries and poor producers in rural areas and benefiting food importers and consumers. The problem, Food Price Watch warns, is that developing countries might have now limited resources to protect vulnerable populations following the economic crisis and stimulus spending.

5. References:

Alejandra Viveros, N. C. (n.d.). *worldbank*. Retrieved from <https://www.worldbank.org/en/news/press-release/2011/11/01/global-food-prices-remain-high-volatile-affecting-poorest-countries-most>.

Bevans, R. (n.d.). *Scibbr*. Retrieved from <https://www.scribbr.com/statistics/t-test/#:~:text=What%20is%20a%20t%2Dtest,means%20is%20different%20from%20zero>.

María Dolores Ugarte, A. F. (n.d.). *PROBABILITY and STATISTICS with R*. CRC Press .

R.H.Riffenburgh. (2012). *Statistics in Medicine* (Third Edition).