American International University- Bangladesh

INTRODUCTION TO DATA SCIENCE

Finished Project Report

Spring 2022-2023

Project Title: **Crypto and The Block Chain Patent Analysis.**
Section: **B**

Student Name	Student Id
Rasel Mahmud	20-43867-2
Sadia Khanom	20-44005-2
Tazin Alam	20-43848-2
A.M. MUFTARIN BILLAH	19-40813-2

Project Overview:

For this project, we have been assigned to scrap data from webpages, perform preprocessing techniques on them, describe them in the light of descriptive statistics and visualize them using R language.

In our project firstly, we chose crypto data of the top 100 crypto in the cryptocurrency market or the digital currency market. We collected crypto data from famous cryptocurrency exchange market website https://www.coingecko.com/. We Collected the data in 27th April. After Gathering Data from the cryptocurrency exchange market website we do some calculations to predict in the next day which crypto price from the top 100 crypto may rise and in the next day in which crypto the investor may interested most. After that we collect the Blockchain patent ownership information from a website. Patent information can be significant in the crypto market for a few reasons. First, it can provide insights into the development and innovation of new technologies and products within the industry. Second, patent information can also provide clues about which companies are investing in crypto-related research and development. This information can be useful for investors looking to make informed decisions about which companies to invest in. Finally, patent information can also be relevant in legal disputes within the industry. In the event of a patent infringement lawsuit, the information contained in the patents can be used as evidence to support or challenge claims of ownership and infringement. We collected the Patent information from the website https://harrityllp.com/titans-of-technology-blockchain-the-top-companies-in-blockchain-patents-2021/. Data Transformation, Data Reduction, and Data Discretization. We did data pre-processing where it was needed. In Descriptive analysis, we described our data with the help of descriptive methods. In the descriptive analysis, we describe our data in some manner and present it in a meaningful way so that it can be easily understood. To describe a comparison between different things we did the Mean, Median, Mode, Range, Variance, Quartile & Percentile. Lastly, we did data visualization to see and understand as visualizations can more effectively allow the reader to digest information. Graphics can allow users to deliver insights in a much easier fashion than describing through text and can also have a greater impact. Here we tried to visualize almost every aspect of comparison & relation.

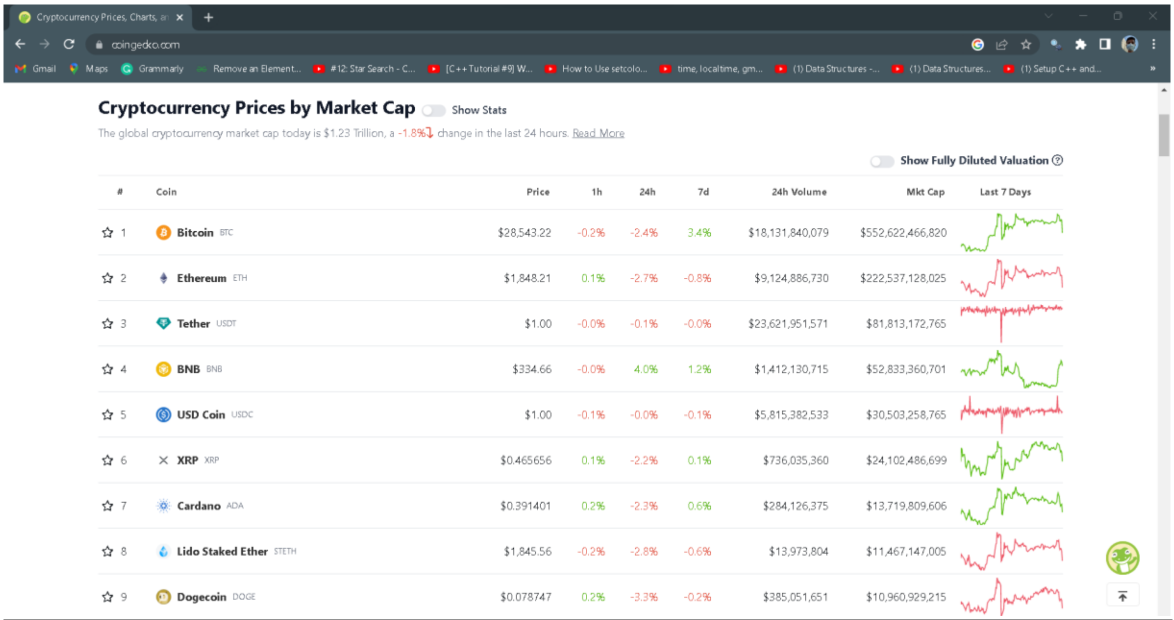
Project Solution Design:

We initially gathered our player lists and performance information for Crypto and and Patent Content from several websites in order to prepare the dataset for data analysis. We then recorded the information in a CSV file. The data pre-processing is then done. Data cleaning is the process of inspecting a raw dataset to find and eliminate errors, duplication, and superfluous data. The table had some missing data, which we replaced with N/A and then filled up with the median. Then we tried to manage every item of noisy data that was in the dataset. After performing data cleaning, measures for data integration, data transformation, data reduction, and data discretization were taken to further clean the data set. We concentrated on using descriptive statistics to rationally simplify our enormous volumes of data after completing the data preprocessing. Moreover, to sum up, the dataset's approximate data. In our data collection, we used the following metrics: Mean, Median, Mode, Range, Variance, Standard Deviation, Quartiles, Percentiles, and Interquartile Ranges. We used data visualization to present facts and data graphically after finishing the descriptive statistics.

Data Collection:

For this project, we start to scrap the data from the website. First, we start to scrap the data from team Coingeko Website for the top 100 crypto market status. In this process, we use a selector gadget to simply select data on a website .

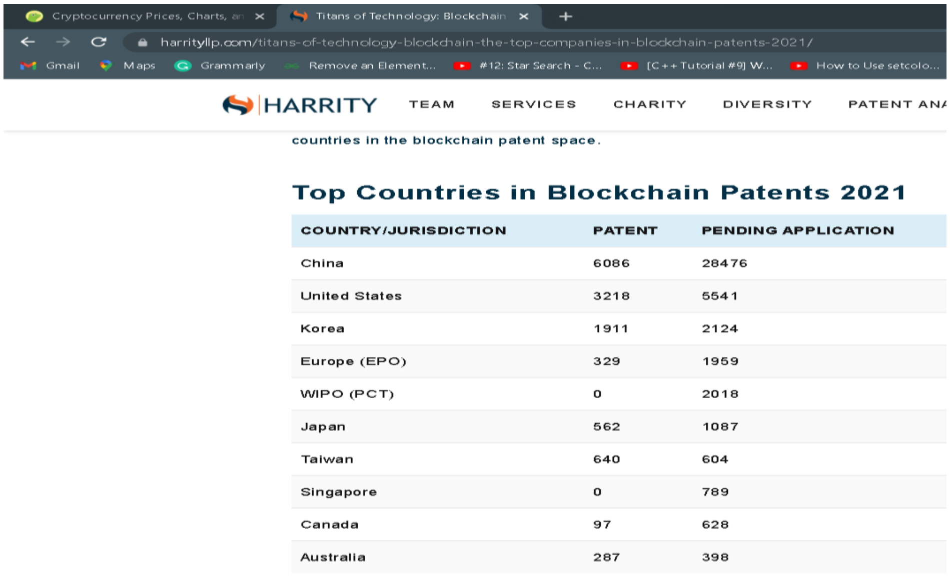
Getting CryptoData:



Code:

```
library(rvest)
library(dplyr)
link="https://www.coingecko.com/"
page=read_html(link)
Coin=page%>%html_nodes(".font-bold")%>%html_text()
Coin <- Coin[-c(101,102)]

Price=page%>%html_nodes(".tw-flex-1 .no-wrap")%>%html_text()
last_1_hour=page%>%html_nodes(".change1h span")%>%html_text()
last_24_hour=page%>%html_nodes(".change24h span")%>%html_text()
last_7_days=page%>%html_nodes(".change7d span")%>%html_text()

last_24hours_volume=page%>%html_nodes(".lit .no-wrap")%>%html_text()
Mkt_Cap=page%>%html_nodes(".cap-price .no-wrap")%>%html_text()

#Data Frame1
Crypto1=data.frame(Coin,Price,last_1_hour,last_24_hour,last_7_days,last_24h
ours_volume,Mkt_Cap,stringsAsFactors = FALSE)
```

Getting Patent Data:



countries in the blockchain patent space.

Top Countries in Blockchain Patents 2021

COUNTRY/JURISDICTION	PATENT	PENDING APPLICATION
China	6086	28476
United States	3218	5541
Korea	1911	2124
Europe (EPO)	329	1959
WIPO (PCT)	0	2018
Japan	562	1087
Taiwan	640	604
Singapore	0	789
Canada	97	628
Australia	287	398

Top Companies in US Blockchain Patents & Per Applications - 2021

Show 25 entries Search:

COMPANY	US PATENTS	US PENDING APPLICATIONS
International Business Machines Corp.	341	435
Advanced New Technologies Co., Ltd.	453	310
Bank Of America Corporation	89	79
Nchain Holdings Limited	3	150
Mastercard Incorporated	45	104
Dell Technologies Inc.	57	53
Capital One Financial Corp.	54	51
Accenture Plc	49	40

Code:

```
> library(rvest)
> library(dplyr)
> response <- read_html("https://harrityllp.com/titans-of-technology-blockc
hain-the-top-companies-in-blockchain-patents-2021/")
> html_text(response)
> tables <- response %>% html_nodes("table") %>% html_table()
> table_one = tables[[1]]
> table_two = tables[[2]]
> table_three = tables[[3]]
> table_four = tables[[4]]
> table_five = tables[[5]]
> table_six = tables[[6]]
> table_seven = tables[[7]]
```

```
> Top_Countries_in_Blockchain_Patents_2021=data.frame(table_one,stringsAsFa
ctors = FALSE)
> Top_Companies_in_Worldwide_Blockchain_Patents_PendingApplications=data.fr
ame(table_two,stringsAsFactors = FALSE)
> Top_Companies_in_US_Blockchain_Patents_Pending_Applications_2021=data.fra
me(table_three,stringsAsFactors = FALSE)
> Top_Companies_in_Chinese_Blockchain_Patents_Pending_Applications_2021=dat
a.frame(table_four,stringsAsFactors = FALSE)
> Top_Companies_Korean_Blockchain_Patents_Pending_Applications_2021=data.fr
ame(table_five,stringsAsFactors = FALSE)
> Top_Companies_in_European_Blockchain_Patents_Pending_Applications_2021=da
ta.frame(table_six,stringsAsFactors = FALSE)
> IBM_vs_Advanced_New_Technologies_Blockchain_Competitive_Gap_Analysis_2021
=data.frame(table_seven,stringsAsFactors = FALSE)
> write.csv(Top_Countries_in_Blockchain_Patents_2021, "Top_Countries_in_Blo
ckchain_Patents_2021.csv", row.names = FALSE)
> write.csv(Top_Companies_in_Worldwide_Blockchain_Patents_PendingApplicatio
ns, "Top_Companies_in_Worldwide_Blockchain_Patents_PendingApplications.csv"
, row.names = FALSE)
> write.csv(Top_Companies_in_US_Blockchain_Patents_Pending_Applications_202
1, "Top_Companies_in_US_Blockchain_Patents_Pending_Applications_2021.csv",
row.names = FALSE)
> write.csv(Top_Companies_in_Chinese_Blockchain_Patents_Pending_Application
s_2021, "Top_Companies_in_Chinese_Blockchain_Patents_Pending_Applications_2
021.csv", row.names = FALSE)
> write.csv(Top_Companies_Korean_Blockchain_Patents_Pending_Applications_20
21, "Top_Companies_Korean_Blockchain_Patents_Pending_Applications_2021.csv"
, row.names = FALSE)
> write.csv(Top_Companies_in_European_Blockchain_Patents_Pending_Applicatio
ns_2021, "Top_Companies_in_European_Blockchain_Patents_Pending_Applications
_2021.csv", row.names = FALSE)
> write.csv(IBM_vs_Advanced_New_Technologies_Blockchain_Competitive_Gap_Ana
lysis_2021, "IBM_vs_Advanced_New_Technologies_Blockchain_Competitive_Gap_An
alysis_2021.csv", row.names = FALSE)
```

Out Put of The Data Frames:

	Coin	Price	last_1_hour	last_24_hour	last_7_days	last_24hours_volume	Mkt_Cap
1	Bitcoin	$28,599.55	0.6%	-2.4%	3.6%	$18,866,677,683	$553,167,905,361
2	Ethereum	$1,850.36	0.2%	-2.9%	-0.7%	$9,342,991,365	$222,618,850,536
3	Tether	$1.00	0.0%	-0.0%	0.0%	$24,747,497,937	$81,775,916,045
4	BNB	$334.04	0.4%	3.5%	1.0%	$1,608,600,977	$52,688,963,340
5	USD Coin	$1.00	0.0%	0.1%	0.1%	$5,961,370,447	$30,524,689,223
6	XRP	$0.464596	-0.2%	-2.6%	-0.2%	$775,830,849	$24,036,701,484
7	Cardano	$0.391849	0.2%	-2.7%	0.7%	$291,501,419	$13,731,059,876
8	Lido Staked Ether	$1,848.32	0.4%	-2.8%	-0.5%	$14,439,420	$11,458,903,794
9	Dogecoin	$0.079189966254	0.3%	-2.8%	0.3%	$394,129,758	$11,005,780,688
10	Polygon	$0.986194	0.3%	-1.6%	-1.8%	$277,880,869	$9,113,587,736
11	Solana	$22.27	0.6%	-5.1%	3.6%	$577,007,243	$8,741,865,230
12	Polkadot	$5.85	0.3%	-2.9%	-1.3%	$141,502,515	$7,178,631,015
13	Litecoin	$87.84	0.4%	-3.3%	1.1%	$429,633,892	$6,395,496,171
14	TRON	$0.068572043590	0.6%	0.9%	2.7%	$336,686,713	$6,220,395,573
15	Binance USD	$1.00	0.4%	0.0%	0.1%	$2,561,632,718	$6,203,212,193
16	Shiba Inu	$0.000010068904	0.3%	-2.2%	-3.0%	$127,252,006	$5,929,237,366
17	Avalanche	$17.10	0.8%	-1.9%	1.3%	$131,207,616	$5,611,564,138
18	Dai	$1.00	-0.0%	0.0%	0.0%	$94,456,409	$4,739,288,951
19	Wrapped Bitcoin	$28,557.21	0.3%	-2.6%	3.6%	$140,360,270	$4,394,676,521

	Country.Jurisdiction	Patent	Pending.Application
1	China	6086	28476
2	United States	3218	5541
3	Korea	1911	2124
4	Europe (EPO)	329	1959
5	WIPO (PCT)	0	2018
6	Japan	562	1087
7	Taiwan	640	604
8	Singapore	0	789
9	Canada	97	628
10	Australia	287	398

	Company	US.Patents	US.Pending.Applications
1	International Business Machines Corp.	341	435
2	Advanced New Technologies Co., Ltd.	453	310
3	Bank Of America Corporation	89	79
4	Nchain Holdings Limited	3	150
5	Mastercard Incorporated	45	104
6	Dell Technologies Inc.	57	53
7	Capital One Financial Corp.	54	51
8	Accenture Plc	49	40
9	Microsoft Corporation	33	54
10	Intel Corporation	26	53
11	Visa Inc.	17	52
12	Toyota Motor Corporation	12	55
13	Salesforce.com, Inc.	12	51
14	Toronto-dominion Bank	26	31
15	Tencent Holdings Ltd	5	52
16	Sony Corporation	15	41
17	Hewlett Packard Enterprise Company	7	48

Data Pre-processing:

Now the most important phase of the data analysis starts which is data pre-processing. We are going to use pre-processing techniques on the datasets to prepare a well completed datasets for analysis and visualization.

1. Data Cleaning

- **Handling Missing Data:** To handle missing data we first need to search the data set for any value that is not assigned. To do so we write a code that will show us the row which contains the missing value,

  Code:

```
missing1 <-
Top_Countries_in_Blockchain_Patents_2021[!complete.cases(Top_Countries_in_Blockchain_Patents_2021),]
print(missing)

missing2 <-
Top_Companies_in_US_Blockchain_Patents_Pending_Applications_2021[!complete.cases(Top_Companies_in_US_Blockchain_Patents_Pending_Applications_2021),]
print(missing)
```

```
missing3 <-
Top_Companies_in_Chinese_Blockchain_Patents_Pending_Applications_2021[!complete.cases(Top_Companies_in_Chinese_Blockchain_Patents_Pending_Applications_2021),]
print(missing)

missing4 <-
Top_Companies_Korean_Blockchain_Patents_Pending_Applications_2021[!complete.cases(Top_Companies_Korean_Blockchain_Patents_Pending_Applications_2021),]
print(missing)

missing5 <-
Top_Companies_in_European_Blockchain_Patents_Pending_Applications_2021[!complete.cases(Top_Companies_in_European_Blockchain_Patents_Pending_Applications_2021),]
print(missing)
```

```
missing <- Crypto1[!complete.cases(Crypto),]
print(missing)
```

Output:

```
> missing1 <- Top_Countries_in_Blockchain_Patents_2021[!complete.cases(Top_Countries_in_Blockchain_Patents_2021),]
> print(missing)
[1] Country.Jurisdiction Patent                Pending.Application
<0 rows> (or 0-length row.names)
> missing2 <- Top_Companies_in_US_Blockchain_Patents_Pending_Applications_2021[!complete.cases(Top_Companies_in_US_Blockchain_P
atents_Pending_Applications_2021),]
> print(missing)
[1] Country.Jurisdiction Patent                Pending.Application
<0 rows> (or 0-length row.names)
> missing3 <- Top_Companies_in_Chinese_Blockchain_Patents_Pending_Applications_2021[!complete.cases(Top_Companies_in_Chinese_Bl
ockchain_Patents_Pending_Applications_2021),]
> print(missing)
[1] Country.Jurisdiction Patent                Pending.Application
<0 rows> (or 0-length row.names)
> missing4 <- Top_Companies_Korean_Blockchain_Patents_Pending_Applications_2021[!complete.cases(Top_Companies_Korean_Blockchain
_Patents_Pending_Applications_2021),]
> print(missing)
[1] Country.Jurisdiction Patent                Pending.Application
```

```
> print(missing)
[1] Country.Jurisdiction Patent                Pending.Application
<0 rows> (or 0-length row.names)
> missing3 <- Top_Companies_in_Chinese_Blockchain_Patents_Pending_Applications_2021[!complete.cases(Top_Companies_in_Chinese_Bl
ockchain_Patents_Pending_Applications_2021),]
> print(missing)
[1] Country.Jurisdiction Patent                Pending.Application
<0 rows> (or 0-length row.names)
> missing4 <- Top_Companies_Korean_Blockchain_Patents_Pending_Applications_2021[!complete.cases(Top_Companies_Korean_Blockchain
_Patents_Pending_Applications_2021),]
> print(missing)
[1] Country.Jurisdiction Patent                Pending.Application
<0 rows> (or 0-length row.names)
> missing5 <- Top_Companies_in_European_Blockchain_Patents_Pending_Applications_2021[!complete.cases(Top_Companies_in_European_
Blockchain_Patents_Pending_Applications_2021),]
> print(missing)
[1] Country.Jurisdiction Patent                Pending.Application
<0 rows> (or 0-length row.names)
```

```
> missing <- Crypto1[!complete.cases(Crypto),]
> print(missing)
[1] Coin              Price             last_1_hour         last_24_hour        last_7_days         last_24hours_volume
[7] Mkt_Cap
<0 rows> (or 0-length row.names)
> |
```

Here We can see that There is no Missing Data in the Data Frames. So Here We don not need to deal with any missing value and we do not need to worry about the missing values.

- **Smooth Noisy Data:** In the dataset, we can see that some columns contain a mixture of both numerical and character data. Like in the Crypto Data Frame Price, Market Cap, Last 24 Volume Contains the contains extra $ sign and last 1h, last 24h and last 7 day the performance contains % as a parameter. For the betterment of the calculation, we have to remove those noises from the dataset.

Code:

In the Crypto Data Frame to remove the $ and , in Price, Market Cap, Last 24 Volume

```
Crypto1$Price <- sub("\\$", "", Crypto1$Price)
Crypto1$Mkt_Cap <- sub("\\$", "", Crypto1$Mkt_Cap)
Crypto1$last_24hours_volume <- sub("\\$", "", Crypto1$last_24hours_volume)

Crypto1$Price <- gsub(",", "", Crypto1$Price)
Crypto1$Mkt_Cap <- gsub(",", "", Crypto1$Mkt_Cap)
Crypto1$last_24hours_volume <- gsub(",", "", Crypto1$last_24hours_volume)
```

In the Crypto Data Frame to remove the % in last 1h, last 24h and last 7 day.

```
Crypto1$last_1_hour <- sub("%", "", Crypto1$last_1_hour)
Crypto1$last_24_hour <- sub("%", "", Crypto1$last_24_hour)
Crypto1$last_7_days <- sub("%", "", Crypto1$last_7_days)
```

Out Put:

	Coin	Price	last_1_hour	last_24_hour	last_7_days	last_24hours_volume	Mkt_Cap
1	Bitcoin	28599.55	0.6	-2.4	3.6	18866677683	553167905361
2	Ethereum	1850.36	0.2	-2.9	-0.7	9342991365	222618850536
3	Tether	1.00	0.0	-0.0	0.0	24747497937	81775916045
4	BNB	334.04	0.4	3.5	1.0	1608600977	52688963340
5	USD Coin	1.00	0.0	0.1	0.1	5961370447	30524689223
6	XRP	0.464596	-0.2	-2.6	-0.2	775830849	24036701484
7	Cardano	0.391849	0.2	-2.7	0.7	291501419	13731059876
8	Lido Staked Ether	1848.32	0.4	-2.8	-0.5	14439420	11458903794
9	Dogecoin	0.079189966254	0.3	-2.8	0.3	394129758	11005780688
10	Polygon	0.986194	0.3	-1.6	-1.8	277880869	9113587736
11	Solana	22.27	0.6	-5.1	3.6	577007243	8741865230
12	Polkadot	5.85	0.3	-2.9	-1.3	141502515	7178631015
13	Litecoin	87.84	0.4	-3.3	1.1	429633892	6395496171
14	TRON	0.068572043590	0.6	0.9	2.7	336686713	6220395573
15	Binance USD	1.00	0.4	0.0	0.1	2561632718	6203212193
16	Shiba Inu	0.000010068904	0.3	-2.2	-3.0	127252006	5929237366
17	Avalanche	17.10	0.8	-1.9	1.3	131207616	5611564138
18	Dai	1.00	-0.0	0.0	0.0	94456409	4739288951
19	Wrapped Bitcoin	28557.21	0.3	-2.6	3.6	140360270	4394676521

There is no need to Smooth rest of the Data Frames.

- **Data Munging:** The dataset does not require munging because all the data are within the same range.

2. Data Reduction:

In our patent holder countries data frames, we want to focus only the top 50 blockchain patent institute/company from every country so we can Reduce the Number of Rows from the data frames.

Code:

```
Top_Companies_in_US_Blockchain_Patents_Pending_Applications_2021 <-
head(Top_Companies_in_US_Blockchain_Patents_Pending_Applications_202
1, 50)
Top_Companies_in_Chinese_Blockchain_Patents_Pending_Applications_202
1 <-
head(Top_Companies_in_Chinese_Blockchain_Patents_Pending_Application
s_2021, 50)
Top_Companies_Korean_Blockchain_Patents_Pending_Applications_2021 <-
head(Top_Companies_Korean_Blockchain_Patents_Pending_Applications_20
21, 50)
Top_Companies_in_European_Blockchain_Patents_Pending_Applications_20
21 <-
```

```
head(Top_Companies_in_European_Blockchain_Patents_Pending_Applicatio
ns_2021, 50)
```

Filter

	Company	US.Patents	US.Pending.Applications
1	International Business Machines Corp.	341	435
2	Advanced New Technologies Co., Ltd.	453	310
3	Bank Of America Corporation	89	79
4	Nchain Holdings Limited	3	150
5	Mastercard Incorporated	45	104
6	Dell Technologies Inc.	57	53
7	Capital One Financial Corp.	54	51
8	Accenture Plc	49	40
9	Microsoft Corporation	33	54
10	Intel Corporation	26	53
11	Visa Inc.	17	52
12	Toyota Motor Corporation	12	55
13	Salesforce.com, Inc.	12	51
14	Toronto-dominion Bank	26	31
15	Tencent Holdings Ltd	5	52
16	Sony Corporation	15	41
17	Hewlett Packard Enterprise Company	7	48
18	Strong Force Tx Portfolio 2018, LLC	1	52
19	AT&T Inc.	26	27

Showing 1 to 20 of 50 entries, 3 total columns

FinalTermProject.R × Crypto1 × FinalProject2.R × Top_Companies_in_Chinese_Blockchain... ×

Filter

	Company	Chinese.Patents	Chinese.Pending.Applications
1	Ping An Insurance (group); Company Of China, Ltd.	107	1768
2	Tencent Holdings Ltd	425	1145
3	Alibaba Group Holding Ltd	310	684
4	Alipay.com Co., Ltd	267	557
5	Shenzhen Oneconnect Technology Co., Ltd.	14	558
6	China United Network Communications Group Company Li...	109	407
7	Hangzhou Fuzamei Technology Co., Ltd.	87	282
8	Baidu, Inc.	79	268
9	China Pingan Property Insurance Stock Co., Ltd.	1	287
10	Shenzhen Qianhai Webank Co., Ltd.	33	209
11	Shenzhen Onething Technology Co., Ltd.	19	204
12	Shandong Aichengshiwang Information Technology Co., Ltd.	0	222
13	Beijing Aimoruice Science And Technology Co., Ltd.	24	195
14	Bank Of China, Ltd.	10	189
15	Jiangsu Rongye Technology Company Limited	9	178
16	Shenzhen Launch Tech Company Limited	7	179
17	Nchain Holdings Limited	1	174
18	Hangzhou Qulian Technology Ltd.	32	121
19	Taikang Life Insurance Co., Ltd	17	128

3. Data Integration:

For the purpose of better analysis, we need to add two Extra Column in Each of the patent holder crypto country name “Country Origin” and “Grand Total Number of Patent”. And in the Crypto Data Frame we will add a Extra Column Name “Track of the Price Percentage” which will hold the summation of price up down percentage from last 7 days to till today.

Patent

	Company	EPO.Patents	EPO.Pending.Applications	Country_Origin	Grand_Total_Number_of_Patent
1	Advanced New Technologies Co., Ltd.	79	294	Europe	373
2	Nchain Holdings Limited	27	153	Europe	180
3	Siemens Ag	23	79	Europe	102
4	Accenture Plc	16	29	Europe	45
5	Nokia Corporation	10	33	Europe	43
6	Alipay.com Co., Ltd	0	40	Europe	40
7	Visa Inc.	4	35	Europe	39
8	Telefonaktiebolaget Lm Ericsson	3	27	Europe	30
9	Mastercard Incorporated	7	23	Europe	30
10	Nec Corporation	5	23	Europe	28
11	Sony Group Corporation	1	26	Europe	27
12	The Government Of Germany	9	18	Europe	27
13	Microsoft Corporation	3	22	Europe	25
14	Huawei Investment & Holding Co., Ltd.	1	23	Europe	24

	Company	Korean.Patents	Korean.Pending.Applications	Country_Origin	Grand_Total_Number_of_Patent
1	Bizmodeline Co Ltd	0	320	Korea	320
2	Alibaba Group Holding Ltd	109	84	Korea	193
3	Coinplug,inc	103	41	Korea	144
4	Nchain Holdings Limited	4	95	Korea	99
5	Samsung Electronics Co., Ltd.	2	57	Korea	59
6	Kt Corporation	11	47	Korea	58
7	Netmarble Corporation	32	24	Korea	56
8	Gold Exchange	0	50	Korea	50
9	Keb Hana Bank	0	47	Korea	47
10	Electronics And Telecommunications Research Institute	6	40	Korea	46
11	Metaps Plus Inc.	44	0	Korea	44
12	Infobank Corp.	0	38	Korea	38
13	Chung-ang University	4	23	Korea	27
14	Sogang University Research & Business Development Foun...	13	12	Korea	25

For the purpose of better analysis, we need to integrate four patent data frames into one complete dataset.

```
> data <- rbind(Top_Companies_in_US_Blockchain_Patents_Pending_Applications
_2021,Top_Companies_in_Chinese_Blockchain_Patents_Pending_Applications_2021
,Top_Companies_Korean_Blockchain_Patents_Pending_Applications_2021,Top_Comp
anies_in_European_Blockchain_Patents_Pending_Applications_2021)
> View(data)
```

	Company	Granted.Patents	Pending.Patents	Country_Origin	Grand_Total_Number_of_Patent
1	International Business Machines Corp.	341	435	USA	776
2	Advanced New Technologies Co., Ltd.	453	310	USA	763
3	Bank Of America Corporation	89	79	USA	168
4	Nchain Holdings Limited	3	150	USA	153
5	Mastercard Incorporated	45	104	USA	149
6	Dell Technologies Inc.	57	53	USA	110
7	Capital One Financial Corp.	54	51	USA	105
8	Accenture Plc	49	40	USA	89
9	Microsoft Corporation	33	54	USA	87
10	Intel Corporation	26	53	USA	79
11	Visa Inc.	17	52	USA	69
12	Toyota Motor Corporation	12	55	USA	67
13	Salesforce.com, Inc.	12	51	USA	63
14	Toronto-dominion Bank	26	31	USA	57
15	Tencent Holdings Ltd	5	52	USA	57
16	Sony Corporation	15	41	USA	56
17	Hewlett Packard Enterprise Company	7	48	USA	55
18	Strong Force Tx Portfolio 2018, LLC	1	52	USA	53
19	AT&T Inc.	26	27	USA	53

Showing 1 to 20 of 200 entries, 5 total columns

A new Column Scale the number of grand total patent in which Grand Total Patent less than 100 or Equal is Scale as Low, Grand Total Patent less than or equal 500 and More than 100 is categorized as Medium Scale, Grand Total Patent less than or equal 1000 and More than 500 is categorized as Medium Large Scale, Grand Total Patent less than or equal 1500 and More than 1000 is categorized as Large Scale and age greater than 1500s is categorized as Excellent Scale.

```
data$Grand_Total_Number_of_Patent <-
as.numeric(data$Grand_Total_Number_of_Patent)
data$Patent_Scale <- case_when(
  data$Grand_Total_Number_of_Patent <= 100 ~ "Low",
  data$Grand_Total_Number_of_Patent <= 500 ~ "Medium",
  data$Grand_Total_Number_of_Patent <= 1500 ~ "Large",
  data$Grand_Total_Number_of_Patent > 1500 ~ "Excellent"
)
```

For a better understanding of the crypto, we integrate a new column named Performance_Progress, which is the sum of the last 1h, last 24h and last 7 days performance.

A new two column named Performance_Progress which is the sum of last 1h, last 24h and last 7 days performance and another One is Growth added in the Crypto1 Data Frame for the better Analysis in Future.

```
Crypto1$Performance_Progress <-
  Crypto1$last_1_hour +
  Crypto1$last_24_hour+Crypto1$last_7_days

Crypto1$Performance_Progress <- as.numeric(Crypto1$Performance_Progress)
Crypto1$Growth <- case_when(
```

```
  Crypto1$Performance_Progress <0 ~ "Negative",
  Crypto1$Performance_Progress ==0 ~ "No Change",
  Crypto1$Performance_Progress >3 ~ "Large Growth",
  Crypto1$Performance_Progress >0 ~ "Positive"
)
```

Out Put:

Performance_Progress	Growth
1.800000e+00	Positive
-3.400000e+00	Negative
0.000000e+00	No Change
4.900000e+00	Large Growth
2.000000e-01	Positive
-3.000000e+00	Negative
-1.800000e+00	Negative
-2.900000e+00	Negative
-2.200000e+00	Negative
-3.100000e+00	Negative
-9.000000e-01	Negative
-3.900000e+00	Negative
-1.800000e+00	Negative
4.200000e+00	Large Growth
5.000000e-01	Positive
-4.900000e+00	Negative
2.000000e-01	Positive
0.000000e+00	No Change
1.300000e+00	Positive

4. Data Transformation

Here we need to convert some of the Data Frames columns values numeric.

Code:

```
#usa
Top_Companies_in_US_Blockchain_Patents_Pending_Applications_2021$US.
Patents <-
as.numeric(Top_Companies_in_US_Blockchain_Patents_Pending_Applicatio
ns_2021$US.Patents)
Top_Companies_in_US_Blockchain_Patents_Pending_Applications_2021$US.
Pending.Applications <-
```

```
as.numeric(Top_Companies_in_US_Blockchain_Patents_Pending_Applicatio
ns_2021$US.Pending.Applications)

#china

Top_Companies_in_Chinese_Blockchain_Patents_Pending_Applications_202
1$Chinese.Patents <-
as.numeric(Top_Companies_in_Chinese_Blockchain_Patents_Pending_Appli
cations_2021$Chinese.Patents)
Top_Companies_in_Chinese_Blockchain_Patents_Pending_Applications_202
1$Chinese.Pending.Applications <-
as.numeric(Top_Companies_in_Chinese_Blockchain_Patents_Pending_Appli
cations_2021$Chinese.Pending.Applications)


#korean

Top_Companies_Korean_Blockchain_Patents_Pending_Applications_2021$Ko
rean.Patents <-
as.numeric(Top_Companies_Korean_Blockchain_Patents_Pending_Applicati
ons_2021$Korean.Patents)
Top_Companies_Korean_Blockchain_Patents_Pending_Applications_2021$Ko
rean.Pending.Applications <-
as.numeric(Top_Companies_Korean_Blockchain_Patents_Pending_Applicati
ons_2021$Korean.Pending.Applications)

#Europian

Top_Companies_in_European_Blockchain_Patents_Pending_Applications_20
21$EPO.Patents <-
as.numeric(Top_Companies_in_European_Blockchain_Patents_Pending_Appl
ications_2021$EPO.Patents)
Top_Companies_in_European_Blockchain_Patents_Pending_Applications_20
21$EPO.Pending.Applications <-
as.numeric(Top_Companies_in_European_Blockchain_Patents_Pending_Appl
ications_2021$EPO.Pending.Applications)
```

We need to change some of the column name in the data frames of the Patent so that we can do the integration between the Data Frames.

Code:

```
colnames(Top_Companies_in_US_Blockchain_Patents_Pending_Applications
_2021)[colnames(Top_Companies_in_US_Blockchain_Patents_Pending_Appli
cations_2021)=="US.Patents"] <- "Granted.Patents"
colnames(Top_Companies_in_US_Blockchain_Patents_Pending_Applications
_2021)[colnames(Top_Companies_in_US_Blockchain_Patents_Pending_Appli
cations_2021)=="US.Pending.Applications"] <- "Pending.Patents"

colnames(Top_Companies_in_Chinese_Blockchain_Patents_Pending_Applica
tions_2021)[colnames(Top_Companies_in_Chinese_Blockchain_Patents_Pen
ding_Applications_2021)=="Chinese.Patents"] <- "Granted.Patents"
colnames(Top_Companies_in_Chinese_Blockchain_Patents_Pending_Applica
tions_2021)[colnames(Top_Companies_in_Chinese_Blockchain_Patents_Pen
ding_Applications_2021)=="Chinese.Pending.Applications"] <-
"Pending.Patents"
```

```
colnames(Top_Companies_Korean_Blockchain_Patents_Pending_Application
s_2021)[colnames(Top_Companies_Korean_Blockchain_Patents_Pending_App
lications_2021)=="Korean.Patents"] <- "Granted.Patents"
colnames(Top_Companies_Korean_Blockchain_Patents_Pending_Application
s_2021)[colnames(Top_Companies_Korean_Blockchain_Patents_Pending_App
lications_2021)=="Korean.Pending.Applications"] <- "Pending.Patents"

colnames(Top_Companies_in_European_Blockchain_Patents_Pending_Applic
ations_2021)[colnames(Top_Companies_in_European_Blockchain_Patents_P
ending_Applications_2021)=="EPO.Patents"] <- "Granted.Patents"
colnames(Top_Companies_in_European_Blockchain_Patents_Pending_Applic
ations_2021)[colnames(Top_Companies_in_European_Blockchain_Patents_P
ending_Applications_2021)=="EPO.Pending.Applications"] <-
"Pending.Patents"
```

	Company	Granted.Patents	Pending.Patents	Country_Origin	Grand_Total_Number_of_Patent
1	Advanced New Technologies Co., Ltd.	79	294	Europe	373
2	Nchain Holdings Limited	27	153	Europe	180
3	Siemens Ag	23	79	Europe	102
4	Accenture Plc	16	29	Europe	45
5	Nokia Corporation	10	33	Europe	43
6	Alipay.com Co., Ltd	0	40	Europe	40
7	Visa Inc.	4	35	Europe	39
8	Telefonaktiebolaget Lm Ericsson	3	27	Europe	30
9	Mastercard Incorporated	7	23	Europe	30
10	Nec Corporation	5	23	Europe	28
11	Sony Group Corporation	1	26	Europe	27
12	The Government Of Germany	9	18	Europe	27
13	Microsoft Corporation	3	22	Europe	25

We have converted the some of Crypto DataFrame column value as numeric for better Analysis.

```
Crypto1$last_1_hour <- as.character(Crypto1$last_1_hour)
Crypto1$last_24_hour <- as.numeric(Crypto1$last_24_hour)
Crypto1$last_7_days <- as.numeric(Crypto1$last_7_days)

Crypto1$Price <- as.numeric(Crypto1$Price)
Crypto1$Mkt_Cap <- as.numeric(Crypto1$Mkt_Cap)
Crypto1$last_24hours_volume <- as.numeric(Crypto1$last_24hours_volume)
```

we need to transform some variables for better analysis of the dataset.

```
Crypto1$Coin <- factor(Crypto1$Coin, ordered = TRUE)
Crypto1$Growth <- factor(Crypto1$Growth,
                            levels =c(1,2,3,4),labels=c("Large
Growth","Positive","No Change","Negative"))

data$Country_Origin <- factor(data$Country_Origin, ordered = TRUE)
data$Patent_Scale <- factor(data$Patent_Scale,
```

```
                                levels
=c(1,2,3,4),labels=c("Excellent","Large","Medium","Low"))
```

5.Data Discretization:

No discretization is needed for this dataset as it is already in a better shape. So we skip this process and move on to descriptive statistics.

Descriptive Statistics:

Now, we are going to compute various descriptive statistics parameters for our dataset.

Firstly, let's try to inspect the central tendency for the various variables of our dataset.

- **MEAN:**

Mean of all top 100 crypto last_1h, last_24h and last_7day market status in Crypto Data Frame.

Code:

```
Meanprice<- mean(Crypto1$Price)
Meanprice
Meanlast_1h_market<- mean(Crypto1$last_1_hour)
Meanlast_1h_market
Meanlast_24_hour<- mean(Crypto1$last_24_hour)
Meanlast_24_hour
Meanlast_7_days<- mean(Crypto1$last_7_days)
Meanlast_7_days
```

```
> Meanprice<- mean(Crypto1$Price)
> Meanprice
[1] 689.2016
> Meanlast_1h_market<- mean(Crypto1$last_1_hour)
> Meanlast_1h_market
[1] 0.289
> Meanlast_24_hour<- mean(Crypto1$last_24_hour)
> Meanlast_24_hour
[1] -1.388
> Meanlast_7_days<- mean(Crypto1$last_7_days)
> Meanlast_7_days
[1] 2.328
>
```

- **MEDIAN:**

Now we calculate the median for the last_24 Market Cap and Volume of the Crypto

Code:

```
medianone=sort(Crypto1$Mkt_Cap)
medianoneres=median(medianone)
medianoneres
mediantwo=sort(Crypto1$last_24hours_volume)
mediantwores=median(mediantwo)
mediantwores
```

OutPut

```
> medianone=sort(Crypto1$Mkt_Cap)
> medianoneres=median(medianone)
> medianoneres
[1] 1003218036
> mediantwo=sort(Crypto1$last_24hours_volume)
> mediantwores=median(mediantwo)
> mediantwores
[1] 39300998
```

- **MODE:**

As the mode doesn't have a built-in function, we first implement the function.

Code:

```
mode <- function(x){
  unique_values <- unique(x)
  table <- tabulate(match(x, unique_values))
  unique_values[table == max(table)]
}
```

```
large_scale_data <- subset(data, Patent_Scale == "Large")
mode_country_origin <- mode(large_scale_data$Country_Origin)
mode_country_origin

excellent_scale_data <- subset(data, Patent_Scale == "Excellent")
mode_country_origin <- mode(excellent_scale_data $Country_Origin)
mode_country_origin

medium_scale_data <- subset(data, Patent_Scale == "Medium")
mode_country_origin <- mode(medium_scale_data $Country_Origin)
mode_country_origin
```

```
small_scale_data <- subset(data, Patent_Scale == " Low ")
mode_country_origin <- mode(small_scale_data $Country_Origin)
mode_country_origin
```

```
> large_scale_data <- subset(data, Patent_Scale == "Large")
> mode_country_origin <- mode(large_scale_data$Country_Origin)
> mode_country_origin
[1] China
```

```
> excellent_scale_data <- subset(data, Patent_Scale == "Excellent")
> mode_country_origin <- mode(excellent_scale_data $Country_Origin)
> mode_country_origin
[1] China
```

```
> medium_scale_data <- subset(data, Patent_Scale == "Medium")
> mode_country_origin <- mode(medium_scale_data $Country_Origin)
> mode_country_origin
[1] China
```

```
> small_scale_data <- subset(data, Patent_Scale == "Low")
> mode_country_origin <- mode(small_scale_data $Country_Origin)
> mode_country_origin
[1] Korea  Europe
```

Range:

Now we calculate the range of variables.

Code:

```
marketvolume=max(Crypto1$last_24hours_volume) - min(Crypto1$last_24hours_volume)
marketvolume
marketcap=max(Crypto1$Mkt_Cap) - min(Crypto1$Mkt_Cap)
marketcap
price=max(Crypto1$Price) - min(Crypto1$Price)
price
grantedpatent=max(data$Granted.Patents) - min(data$Granted.Patents)
grantedpatent
pendingpatent=max(data$Pending.Patents) - min(data$Pending.Patents)
pendingpatent
grandtotalpatent=max(data$Grand_Total_Number_of_Patent) -
min(data$Grand_Total_Number_of_Patent)
grandtotalpatent
```

```
> marketvolume=max(Crypto1$last_24hours_volume) - min(Crypto1$last_24hours_volume)
> marketvolume
[1] 24747497548
> marketcap=max(Crypto1$Mkt_Cap) - min(Crypto1$Mkt_Cap)
> marketcap
[1] 552735095546
> price=max(Crypto1$Price) - min(Crypto1$Price)
> price
[1] 28599.55
>
```

```
> grantedpatent=max(data$Granted.Patents) - min(data$Granted.Patents)
> grantedpatent
[1] 453
> pendingpatent=max(data$Pending.Patents) - min(data$Pending.Patents)
> pendingpatent
[1] 1768
> grandtotalpatent=max(data$Grand_Total_Number_of_Patent) - min(data$Grand_Total_Number_of_Patent)
> grandtotalpatent
[1] 1869
>
```

Quartile & Percentile:

Code:

```
quantile(Crypto1$last_1_hour, prob = c(0.0,0.25,0.50, 0.75 , 0.100))
quantile(Crypto1$last_24_hour, prob = c(0.0,0.25,0.50, 0.75 , 0.100))
quantile(Crypto1$last_7_days)
```

```
> quantile(Crypto1$last_1_hour, prob = c(0.0,0.25,0.50, 0.75 , 0.100))
   0%   25%   50%   75%   10% 
-0.70  0.10  0.40  0.60 -0.01 
> quantile(Crypto1$last_24_hour, prob = c(0.0,0.25,0.50, 0.75 , 0.100))
    0%    25%    50%    75%    10% 
-2.800 -0.825  0.000  1.000 -1.600 
> quantile(Crypto1$last_7_days)
     0%     25%     50%     75%    100% 
-23.000  -6.550  -2.950  -0.175  30.000 
>
```

Interquartile Range:

Code:

IQR(Crypto1$Mkt_Cap)

```
> IQR(Crypto1$Mkt_Cap)
[1] 2205916407
> 
```

Variance:

Code:

var(Crypto1$last_1_hour)
var(Crypto1$last_24_hour)
var(Crypto1$last_7_days)

Output:

```
> var(Crypto1$last_1_hour)
[1] 0.3441455
> var(Crypto1$last_24_hour)
[1] 13.20293
> var(Crypto1$last_7_days)
[1] 43.39929
> 
```

Standard Deviation:

Code:

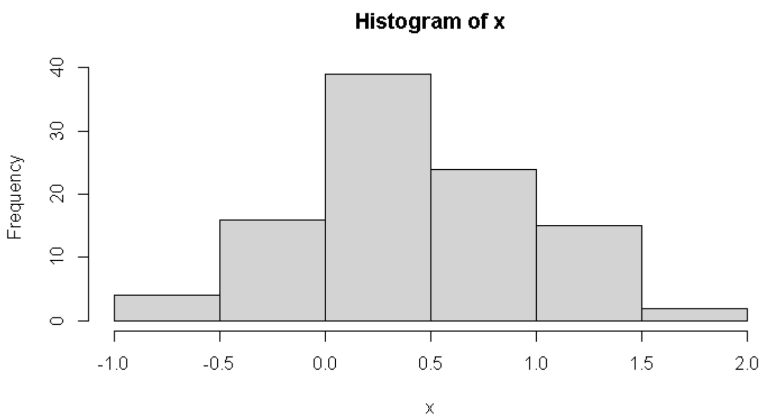
sd(Crypto1$last_1_hour)
sd(Crypto1$last_24_hour)
sd(Crypto1$last_7_days)

```
> sd(Crypto1$last_1_hour)
[1] 0.5866391
> sd(Crypto1$last_24_hour)
[1] 3.633583
> sd(Crypto1$last_7_days)
[1] 6.587814
> 
```

Normal Distribution:

Code:

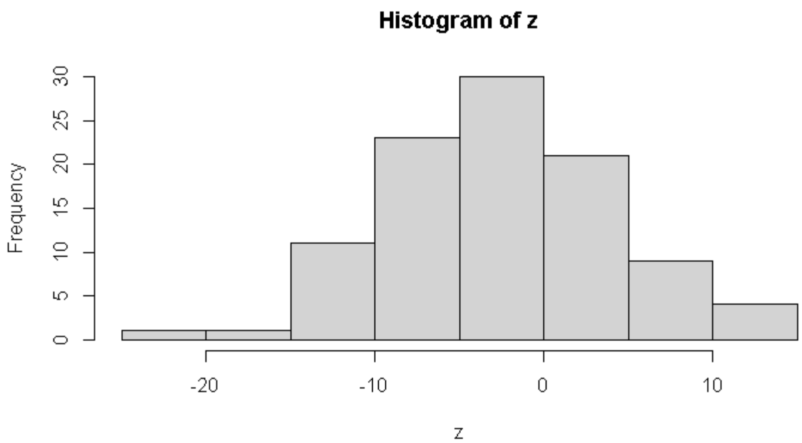
```
x = rnorm(Crypto1$last_1_hour, mean =
mean(Crypto1$last_1_hour), sd=
        sd(Crypto1$last_1_hour))
hist(x)
```



Code:

```
y = rnorm(Crypto1$last_24_hour, mean =
mean(Crypto1$last_24_hour), sd=
        sd(Crypto1$last_24_hour))
hist(y)
```


```
z = rnorm(Crypto1$last_7_days, mean = mean(Crypto1$last_7_days), sd=
        sd(Crypto1$last_7_days))
hist(z)
```

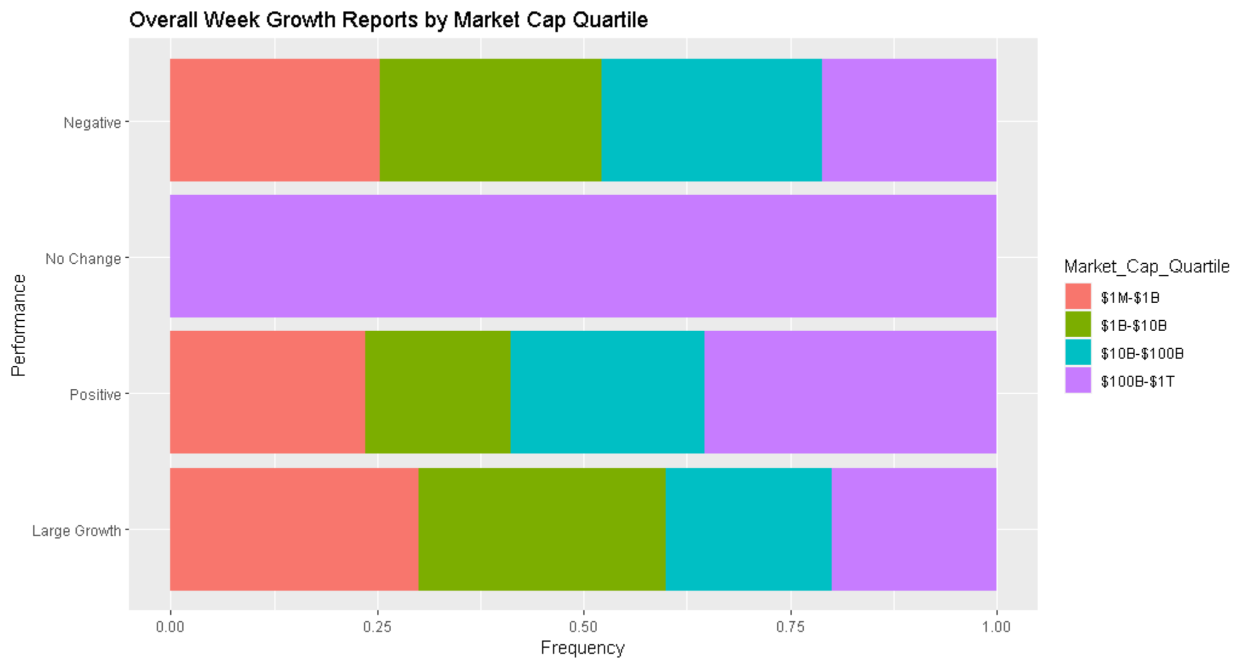


Data Visualization:

1) Now the most important part of the visualization. We need to see the Overall Week Growth of the top 100 Cryptos.

Code:

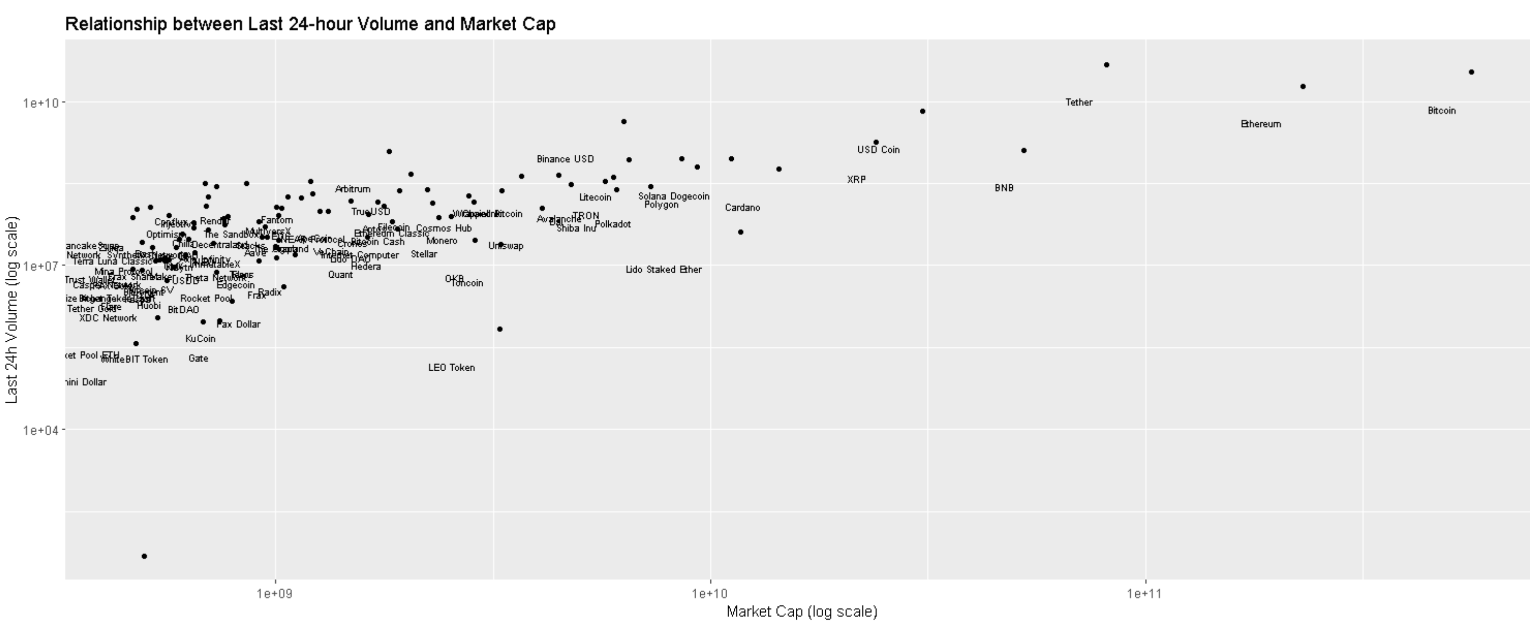
```
library(ggplot2)
Crypto1$Market_Cap_Quartile <- cut(Crypto1$Mkt_Cap, quantile(Crypto1$Mkt_Cap, probs = c(0,
0.25, 0.5, 0.75, 1)), include.lowest = TRUE, labels = c("$1M-$1B", "$1B-$10B", "$10B-$100B",
"$100B-$1T"))

# Create a bar plot of growth by market cap quartile
ggplot(Crypto1, aes(x=Growth, fill=Market_Cap_Quartile)) +
  geom_bar(position = position_fill(reverse = TRUE)) +
  labs(title = "Overall Week Growth Reports by Market Cap Quartile", x = "Performance", y =
"Frequency") +
  coord_flip()
```



This Weekly Report Shows that Most of the coin overall growth remains the same.

2)

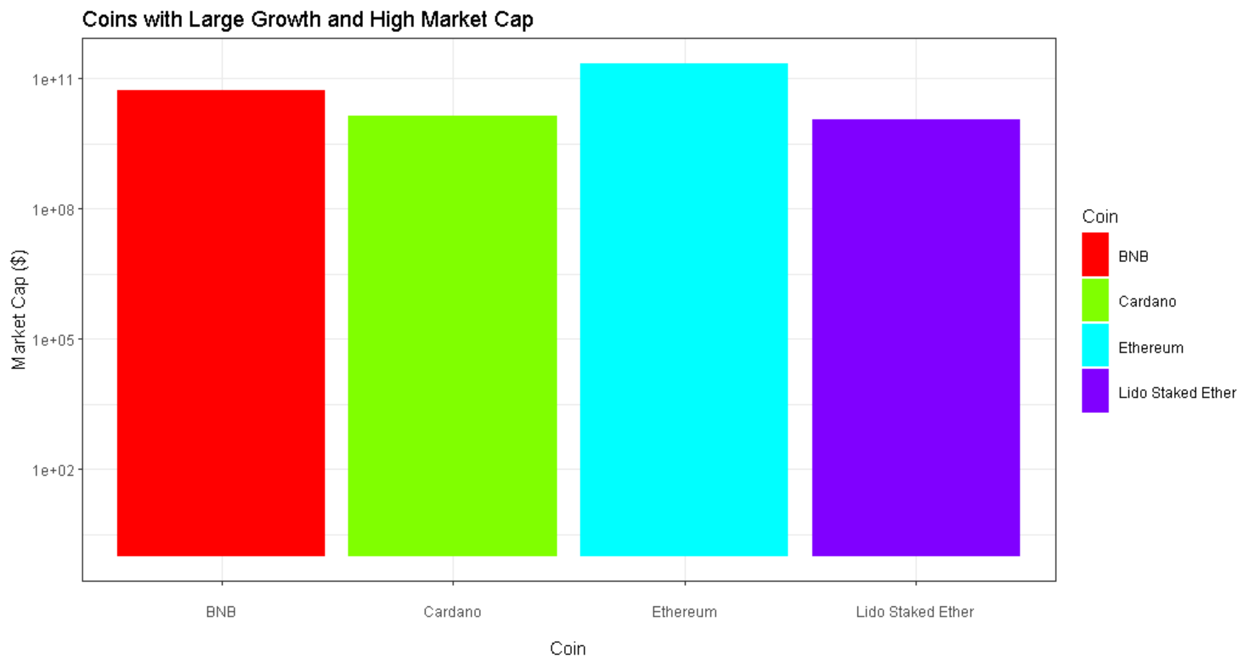
This Scatter Plot Shows the Scatter Plot for Total Market Cap and Last 24 hour volume of the top 100 cryptos in the Coingeko.



The Crypto Which has the Large Market Cap has the Largest Volume in the Last 24 Hour Market Cap. It means the Crypto which has the Largest Market Cap Investor will invest more on these Cryptos as they have a bit lower chance for fall down certainly.

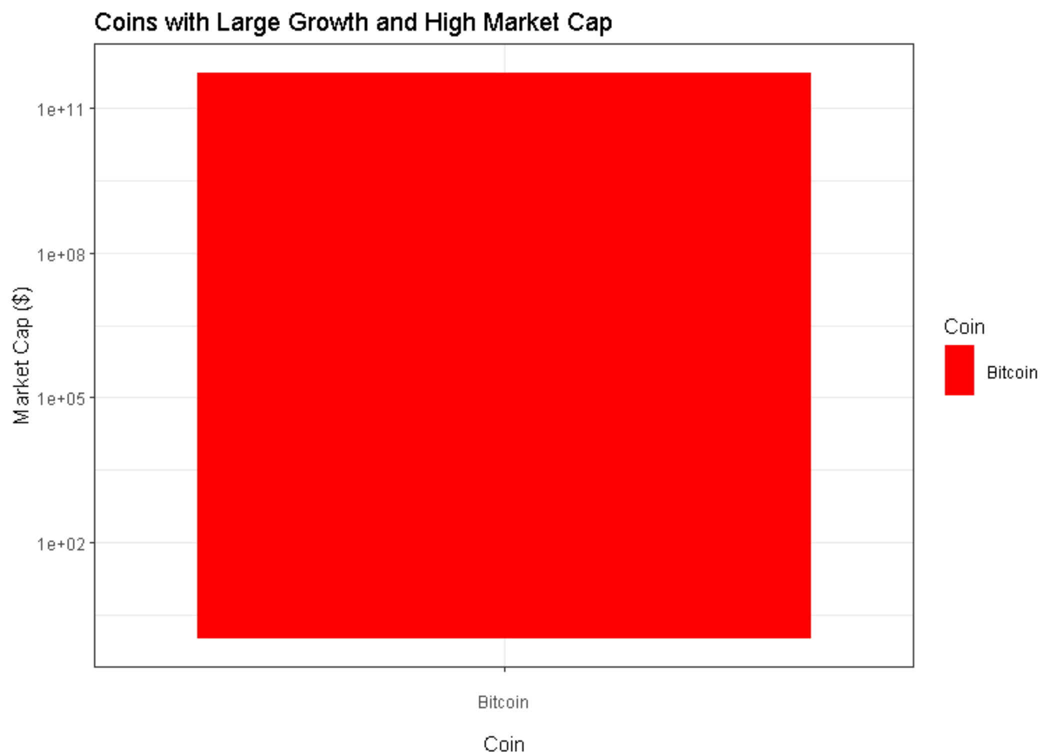
3)

```
filtered_data <- Crypto1 %>%
  filter(Growth == "Positive", Mkt_Cap > 10000000000)

# Create scatter plot of market cap vs. growth for filtered data
ggplot(filtered_data, aes(x = Coin, y = Mkt_Cap, fill = Coin)) +
  geom_bar(stat = "identity", position = "dodge") +
  scale_y_log10() +
  labs(title = "Coins with Large Growth and High Market Cap", x = "Coin", y = "Market Cap ($)") +
  scale_fill_manual(values = rainbow(length(unique(filtered_data$Coin)))) +
  theme_bw()
```



This Graph Shows the Top 4 coin in crypto Market With Large Growth.

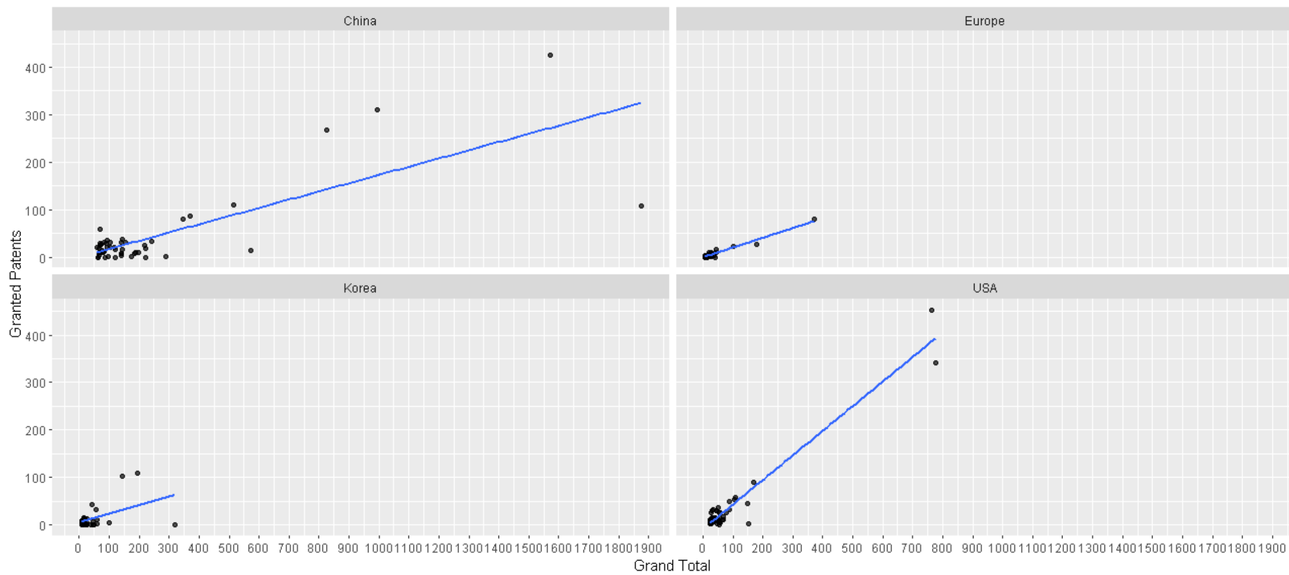
```
filtered_data <- Crypto1 %>%
  filter(Growth == "Large Growth", Mkt_Cap > 10000000000)

# Create scatter plot of market cap vs. growth for filtered data
ggplot(filtered_data, aes(x = Coin, y = Mkt_Cap, fill = Coin)) +
  geom_bar(stat = "identity", position = "dodge") +
  scale_y_log10() +
  labs(title = "Coins with Large Growth and High Market Cap", x = "Coin", y = "Market Cap ($)") +
  scale_fill_manual(values = rainbow(length(unique(filtered_data$Coin)))) +
  theme_bw()
```



This Graph Shows the Top coin in crypto Market With Excellent Growth.

4)

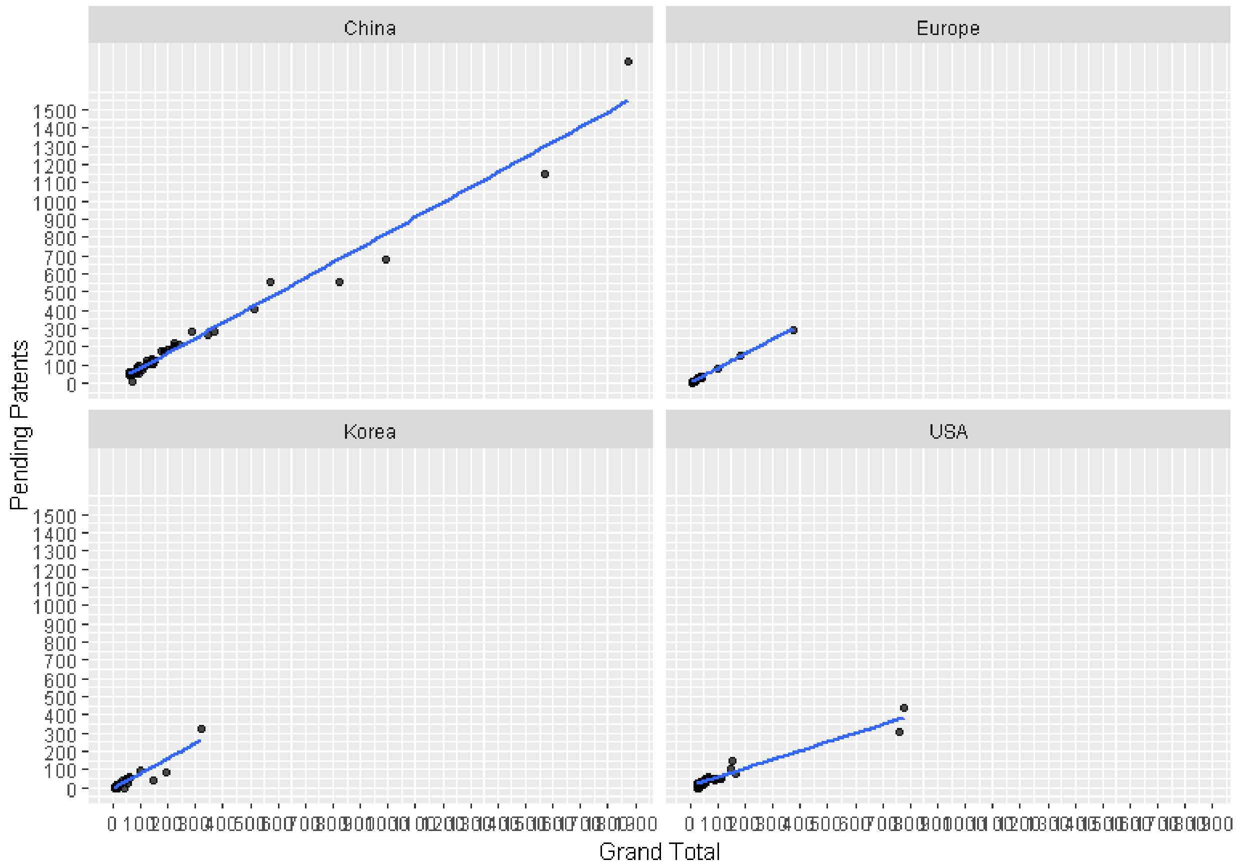
```
ggplot(data, aes(x = Grand.Total, y = Granted.Patents)) +
  geom_point(alpha = 0.7) +
  geom_smooth(method = lm, se = FALSE) +
  scale_x_continuous(breaks = seq(0, 2000, 100)) +
  scale_y_continuous(breaks = seq(0, 1500, 100)) +
  facet_wrap(~ Country_Origin) +
  labs(x = "Grand Total", y = "Pending Patents")
```



This Graph Shows the Country Who have a more focusing in Crypto Market. And China is the Most Granted Patent Holder.

5)

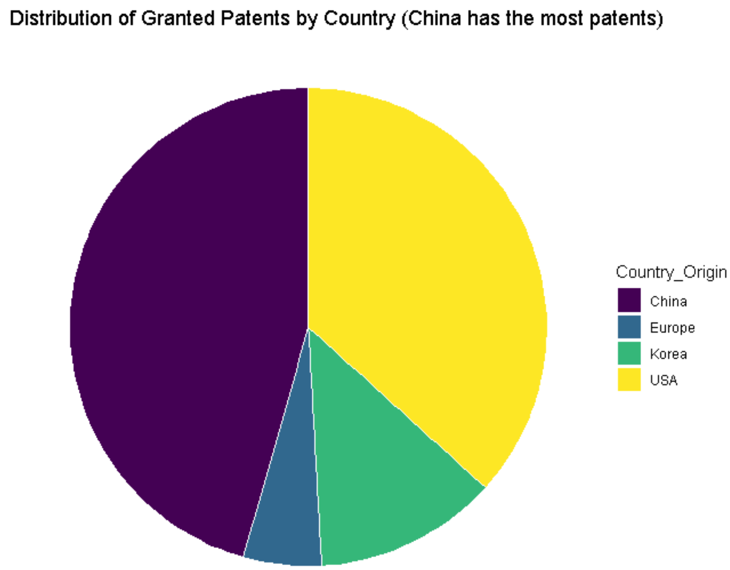
```
ggplot(data, aes(x = Grand.Total, y = Pending.Patents)) +
  geom_point(alpha = 0.7) +
  geom_smooth(method = lm, se = FALSE) +
  scale_x_continuous(breaks = seq(0, 2000, 100)) +
  scale_y_continuous(breaks = seq(0, 1500, 100)) +
  facet_wrap(~ Country_Origin) +
  labs(x = "Grand Total", y = "Pending Patents")
```



This Graph Shows the Country Who have a more focusing in Crypto Market. And China is the Most Pending Patent Holder in the World No Other Country is Even Close to Them.

6)

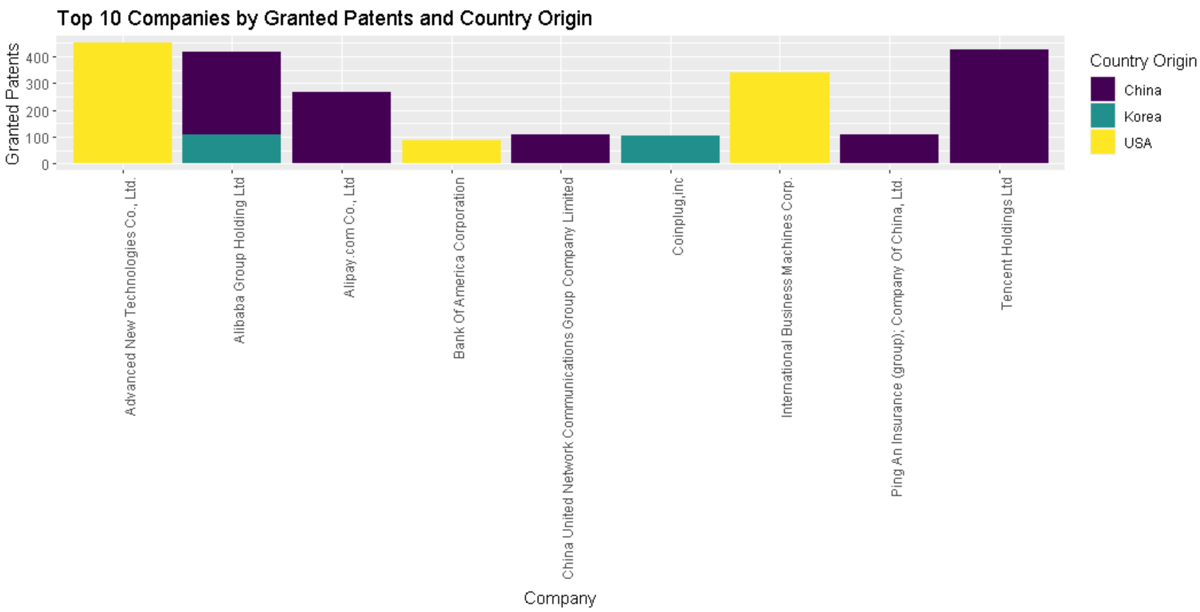
```
df_sum <- data %>%
  group_by(Country_Origin) %>%
  summarize(Total_Patents = sum(Granted.Patents)) %>%
  arrange(desc(Total_Patents))

# Get the country with the maximum granted patents
max_country <- df_sum$Country_Origin[1]

# Create a pie chart with ggplot2
ggplot(df_sum, aes(x = "", y = Total_Patents, fill = Country_Origin)) +
  geom_bar(stat = "identity", width = 1, color = "white") +
  coord_polar(theta = "y") +
  ggtitle(paste0("Distribution of Granted Patents by Country (", max_country, " has the most patents)"))
+
  scale_fill_viridis_d() +
  theme_void()
```



This Graph Shows That China is Holding Almost Worlds 50% Granted Patents in the World.

7)

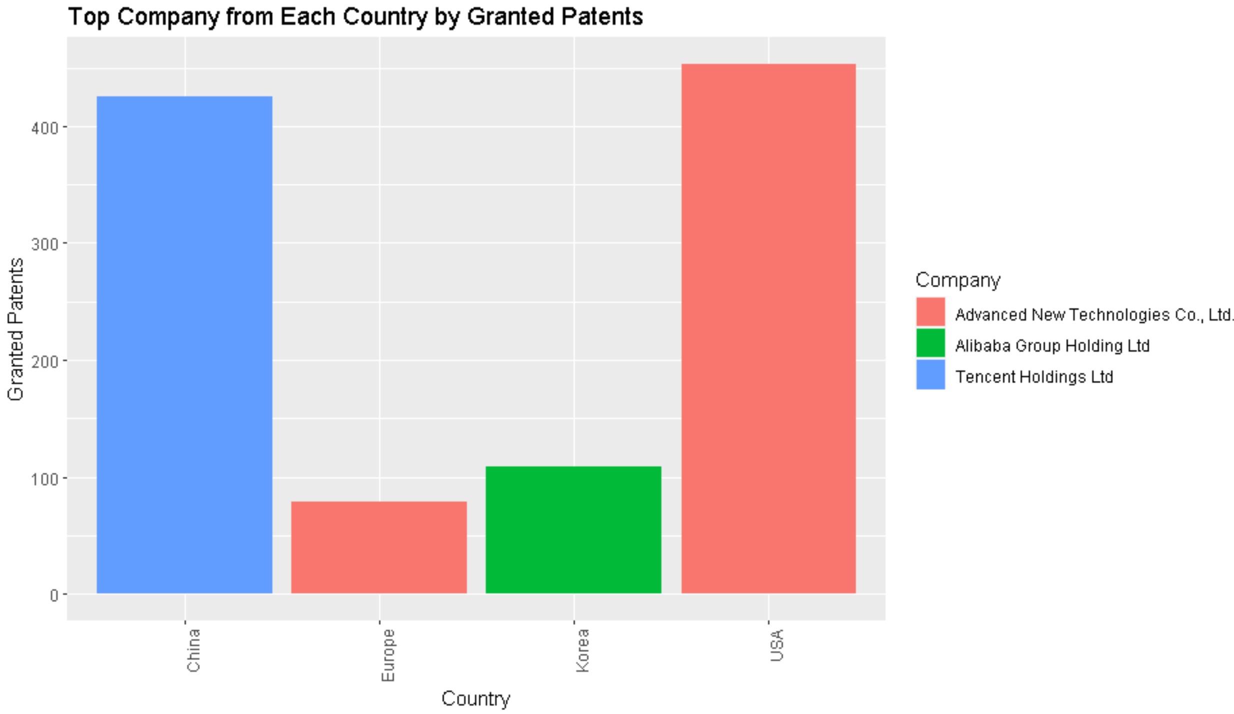
```
data_top <- data %>%
  arrange(desc(Granted.Patents)) %>%
  head(10)

ggplot(data_top, aes(x = Company, y = Granted.Patents, fill = Country_Origin)) +
  geom_bar(stat = "identity") +
  ggtitle("Top 10 Companies by Granted Patents and Country Origin") +
  xlab("Company") +
  ylab("Granted Patents") +
  theme(axis.text.x = element_text(angle = 90, hjust = 1)) +
  guides(fill = guide_legend(title = "Country Origin"))
```



This Graph Shows That

8)

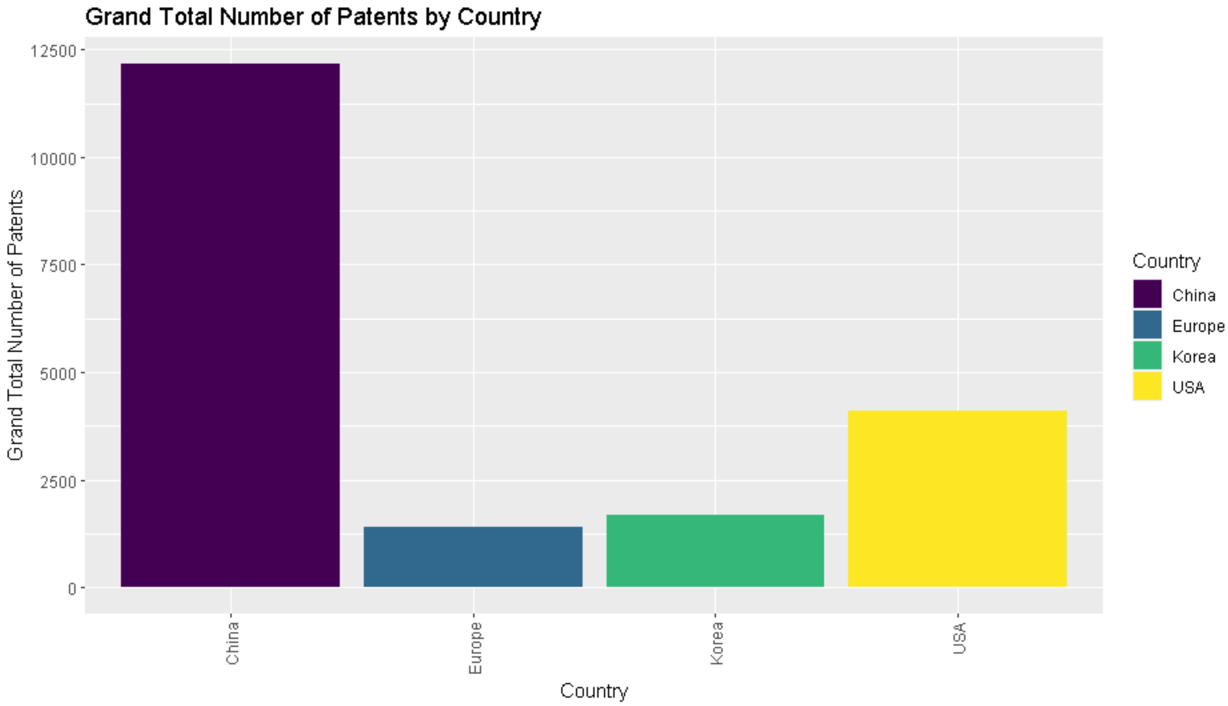
```
# Group data by Country_Origin and get the top company in each group based on Granted.Patents
top_company_by_country <- data %>%
  group_by(Country_Origin) %>%
  slice_max(Granted.Patents) %>%
  ungroup()

# Create a bar chart with ggplot2
ggplot(top_company_by_country, aes(x = Country_Origin, y = Granted.Patents, fill = Company)) +
  geom_bar(stat = "identity") +
  ggtitle("Top Company from Each Country by Granted Patents") +
  xlab("Country") +
  ylab("Granted Patents") +
  theme(axis.text.x = element_text(angle = 90, hjust = 1))
```



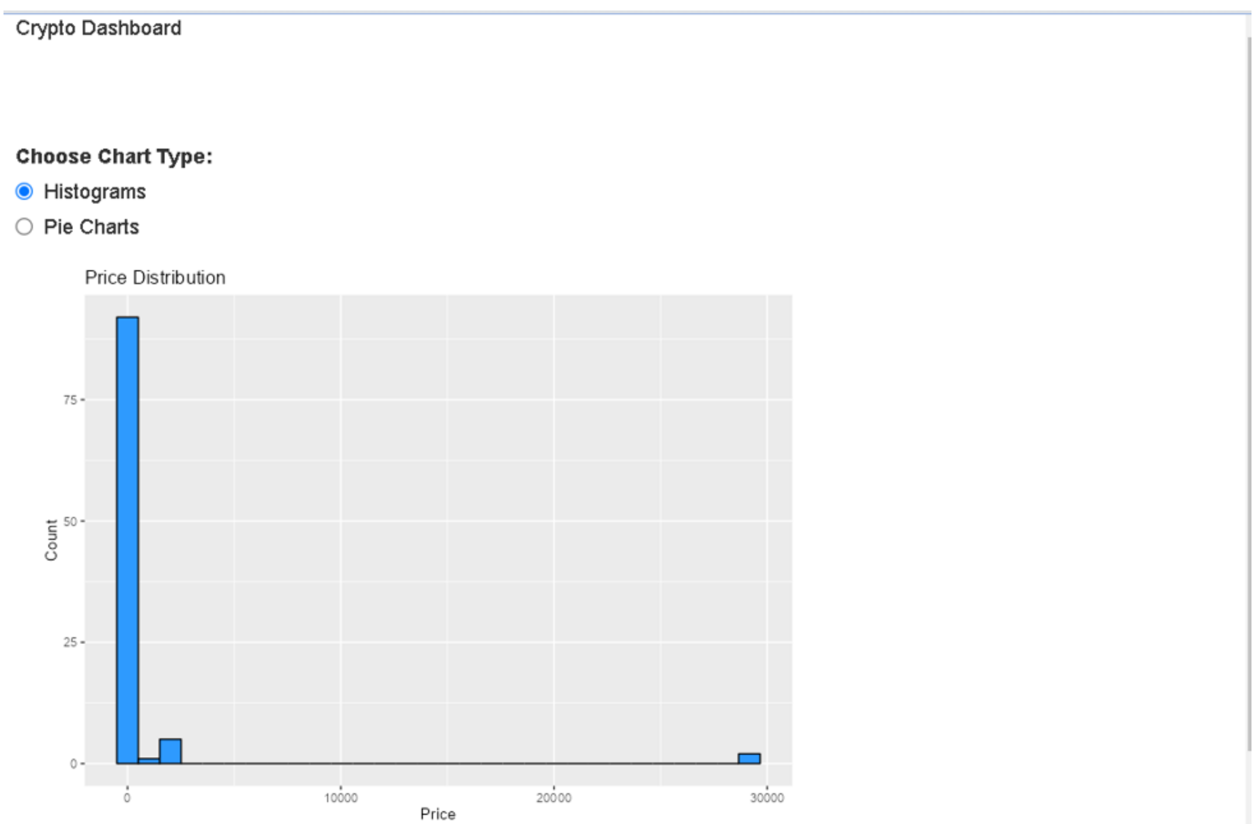
9)

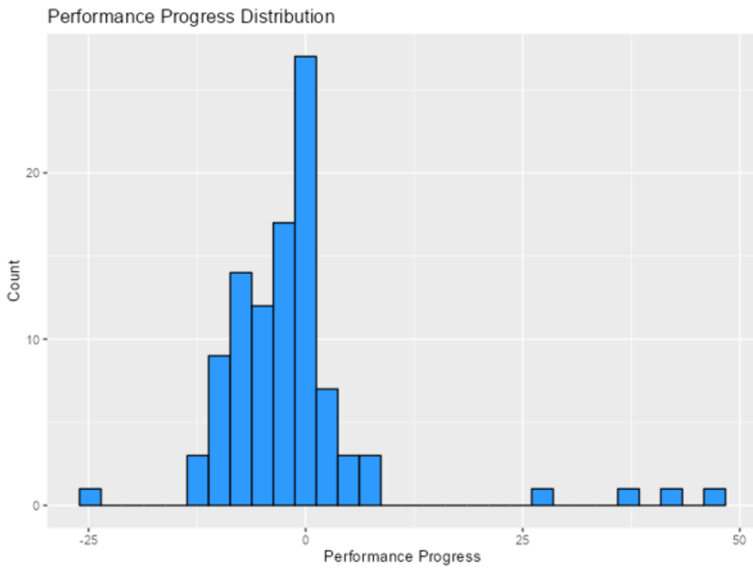
Code:

```
# Sort data by Total_Patents
data_by_country <- data_by_country[order(data_by_country$Total_Patents, decreasing = TRUE), ]

# Create a bar chart
ggplot(data_by_country, aes(x = Country, y = Total_Patents, fill = Country)) +
  geom_bar(stat = "identity") +
  ggtitle("Grand Total Number of Patents by Country") +
  xlab("Country") +
  ylab("Grand Total Number of Patents") +
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1))
```



Shiny Dashboard Implementation:



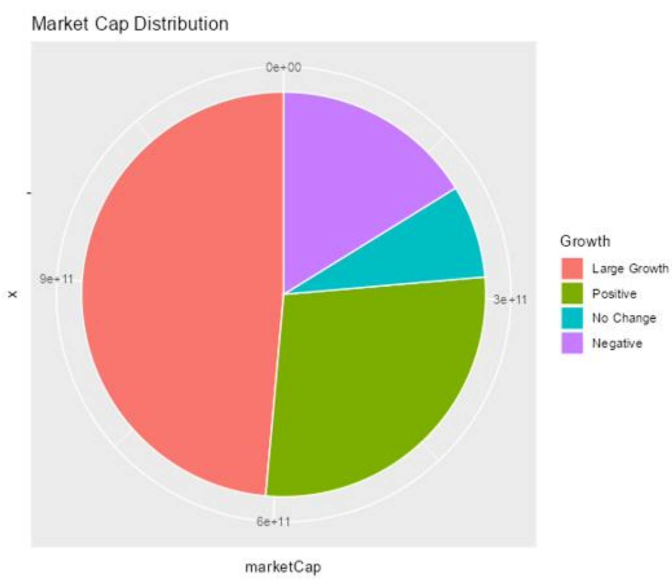


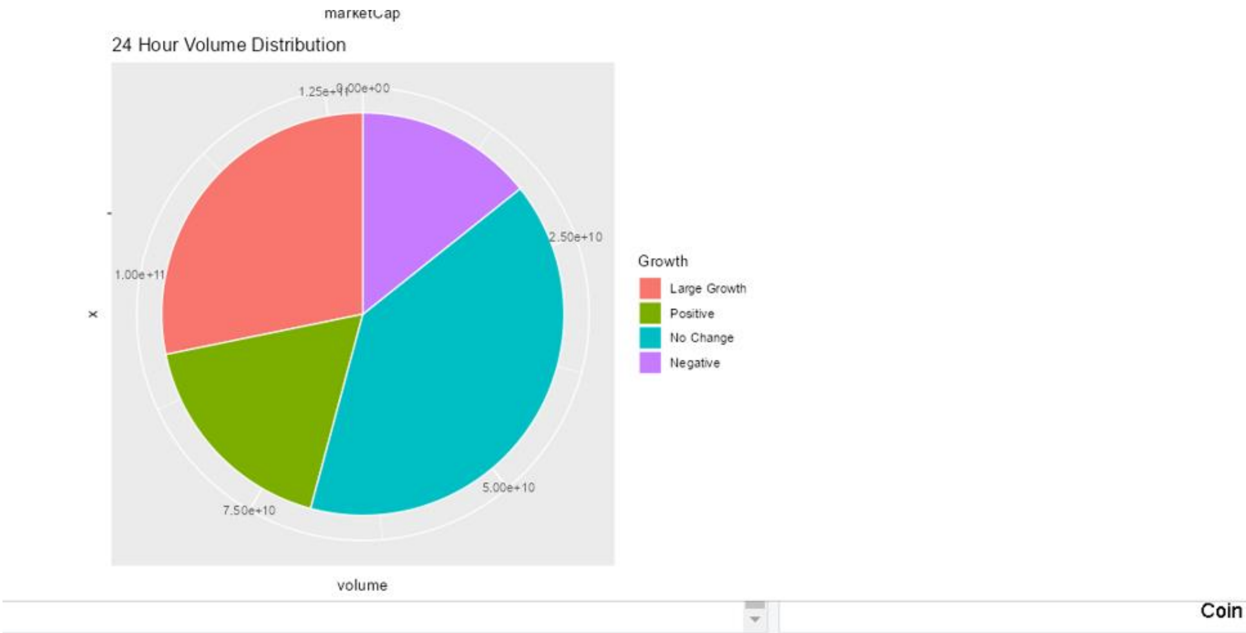
Coin

Crypto Dashboard

Choose Chart Type:

- ○ Histograms
- ◉ Pie Charts





Discussion and Conclusion:

In This Analysis We can see that China will lead the future Blockchain Market. As they have the most granted patent, they have the most pending patent and also they have the most number of total patents. Most of the Top Block Chain Companies are from China. And they almost holds more than 40% of the total Patents of the World. And in the crypto market we can say that investor will invest more in the crypto market in the next day in 2 cryptos they are BNB and Cartano. As they have positive market growth and also they have a good market cap. So IT is a good sign. And market position is also good. And in the last 1h the market price variance is also good it has a low variance. The smaller variance tells that market position is better.