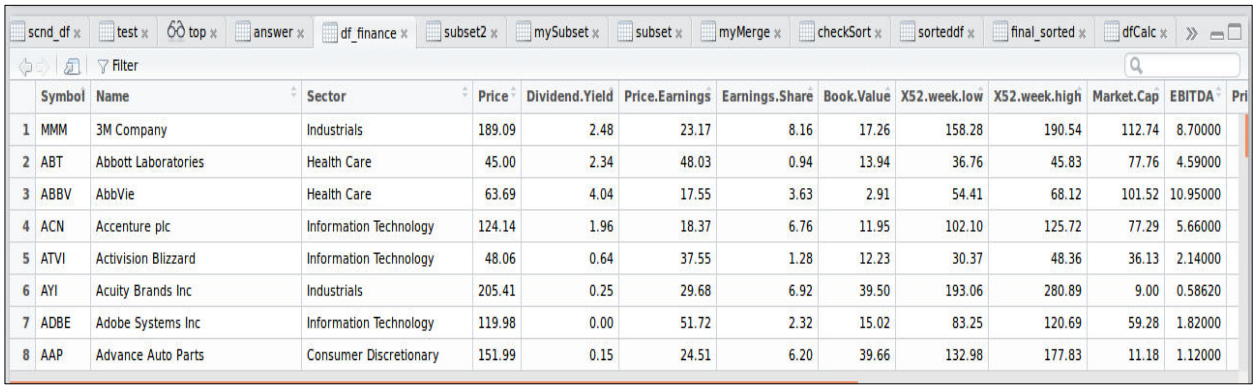
Project 2

1. List all traded corporations found in the dataset.

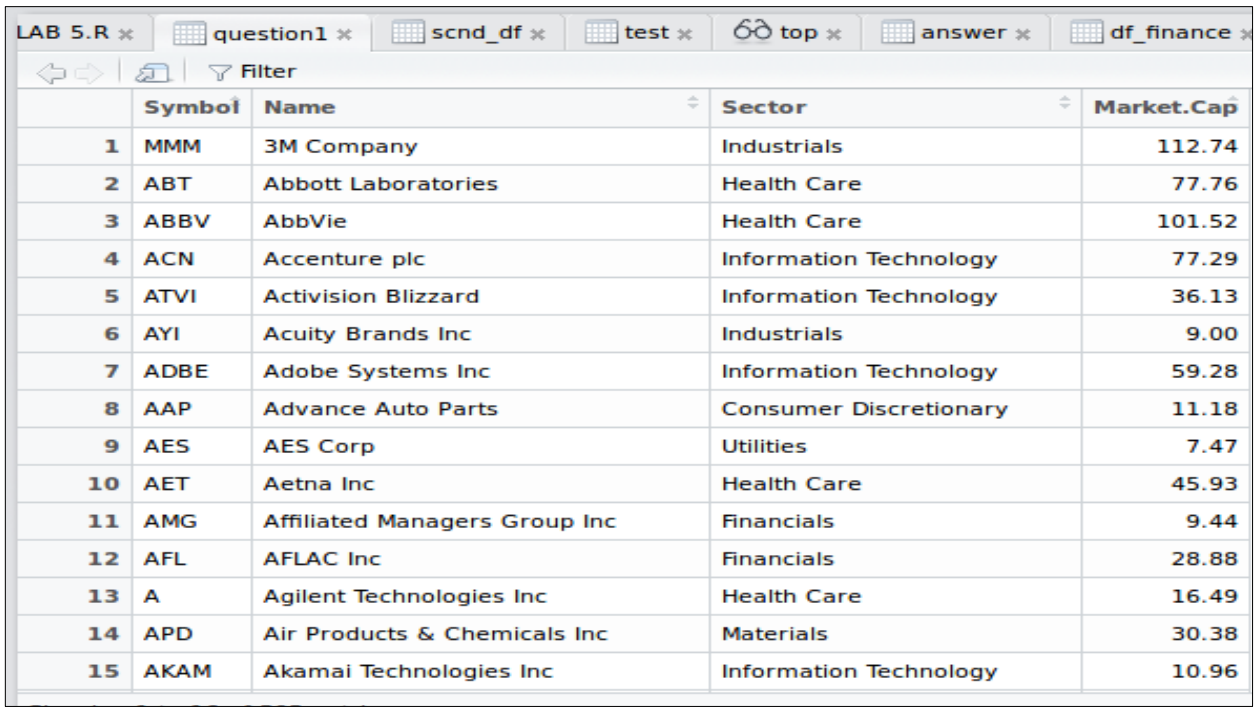
a) List will include 'Num.', 'Symbol Name', 'Security Name', 'Sector', 'Market Cap' sorted by Security Name.

```
> setwd("~/CS-660") //Change work dir
> df_finance <- read.csv("financial_r.csv") //import data
> df_finance //print data frame
```



	Symbol	Name	Sector	Price	Dividend.Yield	Price.Earnings	Earnings.Share	Book.Value	X52.week.low	X52.week.high	Market.Cap	EBITDA	Pri
1	MMM	3M Company	Industrials	189.09	2.48	23.17	8.16	17.26	158.28	190.54	112.74	8.70000	
2	ABT	Abbott Laboratories	Health Care	45.00	2.34	48.03	0.94	13.94	36.76	45.83	77.76	4.59000	
3	ABBV	AbbVie	Health Care	63.69	4.04	17.55	3.63	2.91	54.41	68.12	101.52	10.95000	
4	ACN	Accenture plc	Information Technology	124.14	1.96	18.37	6.76	11.95	102.10	125.72	77.29	5.66000	
5	ATVI	Activision Blizzard	Information Technology	48.06	0.64	37.55	1.28	12.23	30.37	48.36	36.13	2.14000	
6	AYI	Acuity Brands Inc	Industrials	205.41	0.25	29.68	6.92	39.50	193.06	280.89	9.00	0.58620	
7	ADBE	Adobe Systems Inc	Information Technology	119.98	0.00	51.72	2.32	15.02	83.25	120.69	59.28	1.82000	
8	AAP	Advance Auto Parts	Consumer Discretionary	151.99	0.15	24.51	6.20	39.66	132.98	177.83	11.18	1.12000	

```
> question1 <- df_finance[,c("Symbol","Name", "Sector", "Market.Cap")]
> question1[order(question1$Name),] //Order by Security Name
```



	Symbol	Name	Sector	Market.Cap
1	MMM	3M Company	Industrials	112.74
2	ABT	Abbott Laboratories	Health Care	77.76
3	ABBV	AbbVie	Health Care	101.52
4	ACN	Accenture plc	Information Technology	77.29
5	ATVI	Activision Blizzard	Information Technology	36.13
6	AYI	Acuity Brands Inc	Industrials	9.00
7	ADBE	Adobe Systems Inc	Information Technology	59.28
8	AAP	Advance Auto Parts	Consumer Discretionary	11.18
9	AES	AES Corp	Utilities	7.47
10	AET	Aetna Inc	Health Care	45.93
11	AMG	Affiliated Managers Group Inc	Financials	9.44
12	AFL	AFLAC Inc	Financials	28.88
13	A	Agilent Technologies Inc	Health Care	16.49
14	APD	Air Products & Chemicals Inc	Materials	30.38
15	AKAM	Akamai Technologies Inc	Information Technology	10.96

b) List all traded corporations grouped by Sector, ordered by Market Cap (ascending):

```
> qs_b <- dplyr::arrange(question1, Sector, Market.Cap)
> qs_b
```

LAB 5.R | question1 | qs_b | scnd_df | test | top | answer

Filter

	Symbol	Name	Sector	Market.Cap
1	URBN	Urban Outfitters	Consumer Discretionary	2.95
2	SIG	Signet Jewelers	Consumer Discretionary	4.40
3	AN	AutoNation Inc	Consumer Discretionary	4.57
4	TGNA	Tegna, Inc.	Consumer Discretionary	5.54
5	SPLS	Staples Inc.	Consumer Discretionary	5.71
6	BBBY	Bed Bath & Beyond	Consumer Discretionary	5.73
7	KORS	Michael Kors Holdings	Consumer Discretionary	5.84
8	TRIP	TripAdvisor	Consumer Discretionary	5.96
9	RL	Polo Ralph Lauren Corp.	Consumer Discretionary	6.51
10	LEG	Leggett & Platt	Consumer Discretionary	6.63
11	KSS	Kohl's Corp.	Consumer Discretionary	6.87
12	PVH	PVH Corp.	Consumer Discretionary	7.11
13	PHM	Pulte Homes Inc.	Consumer Discretionary	7.24
14	NWSA	News Corp. Class A	Consumer Discretionary	7.36
15	JWN	Nordstrom	Consumer Discretionary	7.49
16	NWS	News Corp. Class B	Consumer Discretionary	7.56
17	HBI	Hanesbrands Inc	Consumer Discretionary	7.57
18	HAR	Harman Int'l Industries	Consumer Discretionary	7.79
19	UA	Under Armour	Consumer Discretionary	7.94

2. For all securities, list their Price-to-Book (P/B) ratio (a valuation multiple useful for value comparison between similar companies within the same industry). Usually, investors look for buying securities with low valuation multiple. Print the 10 ‘min’ and ‘max’ P/B ratios, along with securities and market cap, per sector.

```
> print(df_finance$Price.Book)
```
//List P/B ratio

/* To find 10 Min /Max Securities per Sector I wrote a custom function ***/**

```
> dectTop <- function(x, num, c1,c2){
    sorted<-x[with(x,order(x[,c1],x[,c2],decreasing=F)),]
    splits<-split(sorted,sorted[,c1])
    df<-lapply(splits,head,num)
    do.call(rbind.data.frame,df)}
```
// To find 10 Max Securities per sector.

```
> myMaxSect <- dectTop(test, 10, 3, 5)
```

/* Here, I am calling the function with the following parameters: ***/**

- test: Dataframe
- 10: Number of rows I want to split my result into
- 3: The variable which will be used to group the data (In our case “Sector”)
- 5: The variable which will be used to sort the data (In our case “Price.Book”)

O/P:

	Symbol	Name	Sector	Market.Cap	Price.Book
Consumer Discretion	WYNN	Wynn Resorts Ltd	Consumer Discretion	10.11	64.59
Consumer Discretion	HD	Home Depot	Consumer Discretion	176.1	40.7
Consumer Discretion	AMZN	Amazon.com Inc	Consumer Discretion	403.7	20.94
Consumer Discretion	ORLY	O'Reilly Automotive	Consumer Discretion	24.73	15.34
Consumer Discretion	SBUX	Starbucks Corp.	Consumer Discretion	81.91	14.25
Consumer Discretion	WYN	Wyndham Worldwide	Consumer Discretion	8.63	12.17
Consumer Discretion	ULTA	Ulta Salon Cosmetic	Consumer Discretion	16.8	11.83
Consumer Discretion	TJX	TJX Companies Inc.	Consumer Discretion	51.47	11.46
Consumer Discretion	LOW	Lowe's Cos.	Consumer Discretion	70.01	10.9
Consumer Discretion	ROST	Ross Stores	Consumer Discretion	25.96	9.52
Consumer Staples.11	CLX	The Clorox Company	Consumer Staples	17.56	64.96
Consumer Staples.11	HSY	The Hershey Compa	Consumer Staples	22.97	29.34
Consumer Staples.11	PEP	PepsiCo Inc.	Consumer Staples	156.02	13.92
Consumer Staples.11	K	Kellogg Co.	Consumer Staples	26.03	13.64
Consumer Staples.11	CPB	Campbell Soup	Consumer Staples	18.08	12.27
Consumer Staples.11	MO	Altria Group Inc	Consumer Staples	147.64	11.48
Consumer Staples.11	SYY	Sysco Corp.	Consumer Staples	28.25	11.38
Consumer Staples.11	GIS	General Mills	Consumer Staples	34.95	8.39
Consumer Staples.11	DPS	Dr Pepper Snapple C	Consumer Staples	17.25	8.1
Consumer Staples.10	MNST	Monster Beverage	Consumer Staples	26.46	7.99
Energy.157	OKE	ONEOK	Energy	11.92	63.18
Energy.156	NFX	Newfield Exploration	Energy	7.08	7.62
Energy.155	XEC	Cimarex Energy	Energy	11.6	4.96
Energy.154	HAL	Halliburton Co.	Energy	45.92	4.95
Energy.153	WMB	Williams Cos.	Energy	24.3	4.82
Energy.152	SWN	Southwestern Energy	Energy	3.74	4.26

```
> ascTop <- function(x, num, c1,c2){
    sorted<-x[with(x,order(x[,c1],x[,c2])),]
    splits<-split(sorted,sorted[,c1])
    df<-lapply(splits,head,num)
    do.call(rbind.data.frame,df)}
```

//To find 10 Min Sec, I kept the default value for order()

O/P:

Assignment5ist687.R | IST 687- LAB 5.R | question1 | scnd_df | test | worst | worst2 | top

Filter

	Symbol	Name	Sector	Market.Cap	Price.Book
Consumer Discretionary.1	NWSA	News Corp. Class A	Consumer Discretionary	7.36	0.69
Consumer Discretionary.2	NWS	News Corp. Class B	Consumer Discretionary	7.56	0.70
Consumer Discretionary.3	SPLS	Staples Inc.	Consumer Discretionary	5.71	1.24
Consumer Discretionary.4	GM	General Motors	Consumer Discretionary	56.20	1.30
Consumer Discretionary.5	KSS	Kohl's Corp.	Consumer Discretionary	6.87	1.32
Consumer Discretionary.6	PVH	PVH Corp.	Consumer Discretionary	7.11	1.50
Consumer Discretionary.7	PHM	Pulte Homes Inc.	Consumer Discretionary	7.24	1.54
Consumer Discretionary.8	LEN	Lennar Corp.	Consumer Discretionary	11.89	1.67
Consumer Discretionary.9	F	Ford Motor	Consumer Discretionary	49.52	1.71
Consumer Discretionary.10	DHI	D. R. Horton	Consumer Discretionary	12.27	1.73
Consumer Staples.87	COTY	Coty, Inc	Consumer Staples	14.05	1.44
Consumer Staples.88	ADM	Archer-Daniels-Midland Co	Consumer Staples	25.39	1.49
Consumer Staples.89	TAP	Molson Coors Brewing Company	Consumer Staples	20.85	1.85
Consumer Staples.90	KHC	Kraft Heinz Co	Consumer Staples	110.52	1.93
Consumer Staples.91	SJM	JM Smucker	Consumer Staples	16.21	2.25
Consumer Staples.92	CVS	CVS Health	Consumer Staples	85.74	2.33
Consumer Staples.93	TSN	Tyson Foods	Consumer Staples	22.72	2.38
Consumer Staples.94	MDLZ	Mondelez International	Consumer Staples	66.00	2.65
Consumer Staples.95	WMT	Wal-Mart Stores	Consumer Staples	214.15	2.75

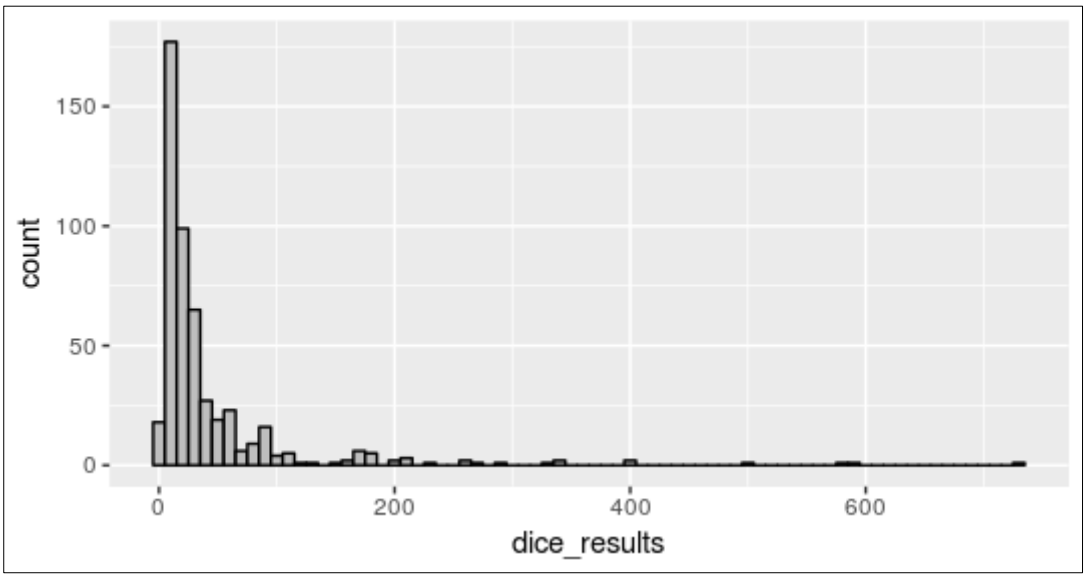
Showing 1 to 19 of 105 entries

3. Plot a histogram (ggplot2) showing securities Market Cap values vs their frequencies. (Extra Points: Include/show sectors)

Histogram plot:

```
> dice_results <- df_finance$Market.Cap
> ggplot() + aes(dice_results)+ geom_histogram(binwidth=10,
colour="black", fill="grey")
```

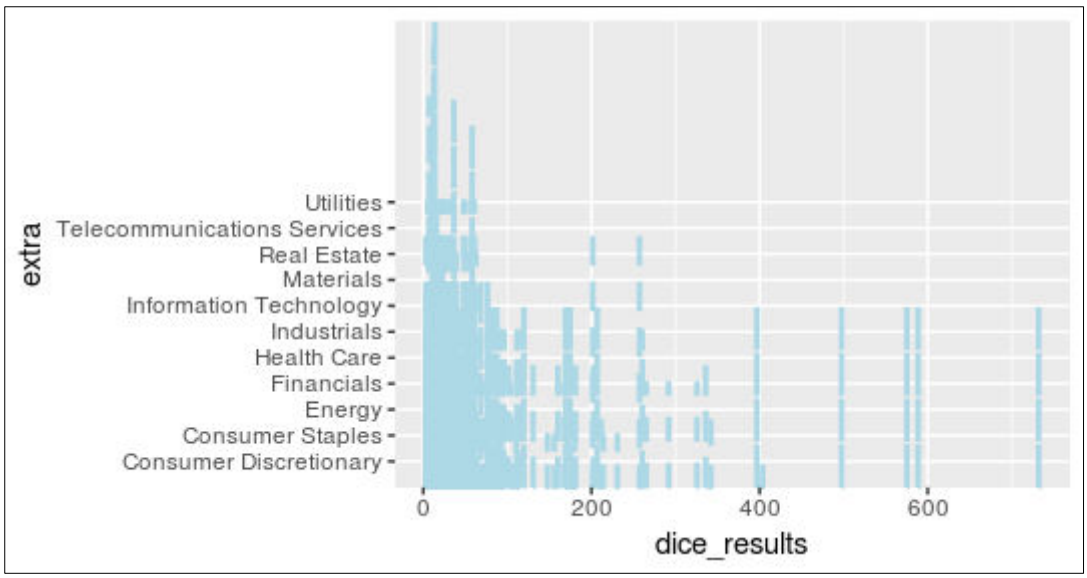
O/P:



Plot with Sector:

```
> extra <- df_finance$Sector
> ggplot() + aes(x=dice_results, y=extra)+ geom_histogram(binwidth=10,
colour="lightblue", fill="blue", stat = "identity", size=1,
linetype="dashed")
```

O/P:



4. From step #2, find the best 5 and worst 5 performed securities (best=low P/B, worst=high P/B). Assume we invested (distributed equivalently) $ 10,000 on all 10 securities what will be our income just on dividends (if any)?

```
> checkSort <- dplyr::arrange(df_finance,desc(df_finance$Price.Book))
```
//Sort via Price.Book
```
> myHead <- head(checkSort, n=5)
> myTail <- tail(checkSort, n=5)
> myMerge <- rbind(myHead, myTail)
```
//Best 5 and Worst 5 Securities based on P.B ratio
```
> myMerge["Divided_Based_Income"] <- myMerge$Dividend.Yield*1000
```
//Added new column to show income results

O/P:

myMerge

Filter

	Symbol	Name	Sector	Price	Dividend.Yield	Earnings.Share	Book.Value	X52.week.low	X52.week.high	Market.Cap	Price.Book	Divided_Based_Income
1	UPS	United Parcel Service	Industrials	105.64	3.13	3.87	0.47	98.85	120.44	91.91	227.68	3130
2	BA	Boeing Company	Industrials	182.02	3.12	7.61	1.32	121.43	185.71	111.48	136.65	3120
3	CLX	The Clorox Company	Consumer Staples	136.91	2.34	4.99	2.11	111.24	140.47	17.56	64.96	2340
4	WYNN	Wynn Resorts Ltd	Consumer Discretionary	99.46	1.98	2.38	1.55	79.77	109.50	10.11	64.59	1980
5	OKE	ONEOK	Energy	56.58	4.36	1.66	0.90	25.61	59.47	11.92	63.18	4360
480	FTR	Frontier Communications	Telecommunications Services	2.62	14.63	-0.51	3.85	2.57	5.75	3.07	0.72	14630
481	NWS	News Corp. Class B	Consumer Discretionary	13.00	1.51	-0.63	18.84	10.90	15.22	7.56	0.70	1510
482	NWSA	News Corp. Class A	Consumer Discretionary	12.66	1.55	-0.63	18.84	10.54	14.68	7.36	0.69	1550
483	FSLR	First Solar Inc	Information Technology	32.63	NA	-3.48	50.10	28.60	74.29	3.39	0.67	NA
484	RIG	Transocean	Energy	12.78	0.00	2.10	41.27	8.34	16.66	4.67	0.32	0

5. Should we have invested 40% on AAPL, 30% on AMZN, and 30% on GOOGL securities what would have been our total (relative) return (assume we bought at '52 week low' and sold at '52 week high' --- assume the age-old strategy "Buy low and sell high" is valid!!)

```
> mySubset <- df_finance[df_finance$Symbol == "GOOGL" |
df_finance$Symbol == "AMZN" | df_finance$Symbol == "AAPL",]
```
//Extracted only GOOGL, AMZN, AAPL data

// Performing relative return operation i.e. (52WeekHigh - 52WeekLow) / 52WeekLow

```
> relative <- ((mySubset[1:n,10] - mySubset[1:n,9]) / mySubset[1:n,9])
```
// This gives numerical values specifying the Relative returns for respective securities.

```
> mySubset["Relative_Return"] <- relative
```
//Added new variable for Relative_Return

O/P:

myMerge | mySubset

Filter

	Symbol	Name	Sector	Price	Dividend.Yield	Earnings.Share	Book.Value	X52.week.low	X52.week.high	Market.Cap	Price.Book	Relative_Return
24	GOOGL	Alphabet Inc Class A	Information Technology	851.15	NA	27.88	201.12	672.66	867.00	588.5	4.21	0.2889127
27	AMZN	Amazon.com Inc	Consumer Discretionary	846.02	NA	4.90	40.43	538.58	860.86	403.7	20.94	0.5983884
46	AAPL	Apple Inc.	Information Technology	139.52	1.63	8.33	25.19	89.47	140.28	732.0	5.53	0.5678999

Showing 1 to 3 of 3 entries