VIETNAM NATIONAL UNIVERSITY HO CHI MINH CITY
UNIVERSITY OF ECONOMICS AND LAW

MIDTERM REPORT

Subject: PROGRAM PACKAGE IN FINANCE 2

LECTURER: PhD. Nguyen Thanh Liem

Submitted by: Nguyen Thi Hue Minh
ID: K194141733

Ho Chi Minh City, May 10th, 2022

Contents

1. Literature review

The amount of cash reflects a company's liquidity, and asset structure. The amount of cash shows whether the manager is taking advantage of resources to carry out production and business activities effectively or not. A company with a large amount of cash may be in a state of excess, underutilizing its resources. If cash is low, the company may find it difficult to pay its debts and other expenses. In addition, the amount of cash held affects the ability of the company to handle unexpected difficulties.

$$Cash\ holding = \frac{Cash}{Total\ asset}$$

Leverage is continuous variable and firm size is discre variable.

Leverage: The trade-off theory and pecking order theory and the free cash flow theory suggest that there is a negative relationship between leverage and cash holding. According to Ozkan and Ozkan (2004), and Diamond (1984), the trade-off theory holds that debt reduces moral hazard so it can replace cash. The company's use of less debt allows managers to decide to hold more cash. According to Chen and Liu (2003), Saddour (2006) also shows that leverage and cash holding are negatively related. When the company has a large amount of cash, the use of external funding sources will be less.

$$Leverage = \frac{Total\ debt}{Total\ asset}$$

Firm size: The trade-off theory suggests that there is a negative relationship between firm size and cash holding because large firms have easy access to cheap sources of financing (Ferri and Jones, 1979). In contrast, pecking order theory suggests that firm size and cash holdings are positively related because the costs of large firms are higher than those of small firms that require a lot of cash. Kafayat and et al (2014) also show that there is a positive correlation between firm size and cash holding. As the company expands, operating costs and research investments must also increase, so the amount of cash needs to increase to pay.

Firm size is generated from calculating the nature logarithm of the total asset. Firms with the natural logarithm of total assets greater than the median of all these values are classified as large, and conversely, as small.

$$\text{ln\_asset} = \ln(Total\ asset)$$

- ln_asset > 0: size = “Big”
- ln_asset <0 : size = “Small”

In this report, the hypothesis proposed in this study is as follows:

- Leverage and cash holding have a negative correlation.
- Firm size and cash holding have a positive correlation.

2. Create dataset

2.1. Create random sample

Use sample() to generate a random sample of 100 random firms from the number of firms in the provided data set. The set.seed() function is used to get the same result using sample(). Use the generated random sample to generate a new dataset.

```
set.seed(733)
sample = sample(1:nrow(data), 100, replace = F)
df = data[sample,]
```

Calculate cash holding by dividing cash by total assets:

```
df$cash_holding = df$cash/df$totalasset
```

Calculate leverage (continuous variable) by dividing total debt by total asset:

```
df$leverage = df$totaldebt/df$totalasset
```

2.2. Handle NA value and create discreate variable

Check missing value of variables. Because total asset is used to create discrete varible, check its missing value. Total asset have no missing value. Leverage and cash holding has 4 missing value. Replace missing value of leverage/cash holding by median of leverage/cash holding.

```
## Transform NA
sum(is.na(df$leverage)) #1
sum(is.na(df$totalasset)) #0
sum(is.na(df$cash_holding)) #4

df$leverage[is.na(df$leverage)] = median(df$leverage, na.rm=T)
df$cash_holding[is.na(df$cash_holding)]= median(df$cash_holding,na.rm=T)
```

Create size (discrete variable) by nature logarithm (base e) of the total asset with 3 values: Small, Big. Create ln_asset column with nature logarithm of total asset: values less than median of ln_asset are “Small”, values larger than ln_asset are “Big”.

```
df$ln_asset= log(df$totalasset)

df$size = 0

df$size[df$ln_asset<= median(df$ln_asset)] = "Small"
df$size[df$ln_asset>= median(df$ln_asset)] ="Big"
```

Count the number of firms each value of the size variable:

```
df %>%
  group_by(size) %>%
  summarise(count=n())
```

```
  size  count
  <chr> <int>
1 Big      50
2 Small    50
```

The generated data has 50 firms of Big and 50 firms of Small.

3. Report

3.1. Find 5 firm with highest cash holding

I use arrange() to rearrange the cash holding values in order, with desc being the descending value. Use select() to represent the company and its respective industry. Use slice() to select the first 5 values of the cash holding after reordering.

```
print('5 firms with highest cash holding are: ')
df %>% arrange(desc(cash_holding)) %>%
  select(firmname, cash_holding) %>%
  slice(1:5)
```

```
  firmname                                cash_holding
  <chr>                                          <dbl>
1 Saigon Beer Alcohol Beverage Corp              0.676
2 Pharmedic Pharmaceutical Medicinal JSC         0.601
3 Vinacafe Bien Hoa JSC                          0.578
4 Clever Group Corp                              0.524
5 Dhg Pharmaceutical Joint-Stock Co              0.466
```

The Saigon Beer Alcohol Beverage Corp has the highest cash holding with 0.676. The second is Pharmedic Pharmaceutical Medicinal JSC with 0.601. Followed by Vinacafe Bien Hoa JSC with 0.578, Clever Group Corp with 0.524, and Dhg Pharmaceutical Joint-Stock Co with 0.466.

3.2. Find 5 firm with lowest cash holding

To find 5 firms with the lowest cash holding, use the same code as above. Use arrange() to rearrange cash holding in ascending order. Keep using select() and slice() to get the first 5 firm name with corresponding cash holding after reordering.

```
df %>%  arrange(cash_holding) %>%
  select(firmname, cash_holding) %>%
  slice(1:5)
```

	firmname	cash_holding
	<chr>	<dbl>
1	Vinacomin Coc Sau Coal JSC	0.000650
2	Vinh Plastic and Bags JSC	0.00103
3	Vexilla Viet Nam Group JSC	0.00157
4	Viet Tien Son Real Estate Holding Co	0.00223
5	Kim Vi Inox Import Export Production JSC	0.00413

The result shows that 5 firm has the lowest cash holding are: Vinacomin Coc Sau Coal JSC, Vinh Plastic and Bags JSC, Vexilla Viet Nam Group JSC, Viet Tien Son Real Estate Holding Co, Kim Vi Inox Import Export Production JSC with correspoding cash holding are 0.00065, 0.00103, 0.00157, 0.00223, 0.00413. In which, Vinacomin Coc Sau Coal JSC has the lowest cash holding with 0.00065.

3.3. The name of industries with firms belong to

Change industry column from continuous to discrete. And use levels() to list industries.

```
df$industry=as.factor(df$industry)
levels(df$industry)
```

```
 [1] "Basic Materials"        "Consumer Cyclicals"     "Consumer Non-Cyclicals"
 [4] "Energy"                 "Financials"             "Healthcare"
 [7] "Industrials"            "Real Estate"            "Technology"
[10] "Utilities"
```

The data has 10 industries including: Basic Materials, Consumer Cyclicals, Consumer Non-Cyclicals, Energy, Financials, Healthcare, Industrials, Real Estate, Technology, Utilities.

3.4. The descriptive statistic

3.4.1. Different categories of the discrete variable

```
df %>%
  group_by(size) %>%
  summarize(min = min(cash_holding),
            max = max(cash_holding),
            median = median(cash_holding),
            mean = mean(cash_holding),
            standard_deviation = sd(cash_holding))
```

size	min	max	median	mean	standard_deviation
<chr>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
Big	0.000650	0.676	0.127	0.178	0.158
Small	0.00103	0.601	0.0592	0.125	0.150

- Big: Min value = 0.000650, among 50 big firms, there is at least 1 company with a cash holding ratio of 0.000650 – this is the lowest cash holding ratio. Max value = 0.676, the highest cash holding ratio in big firms is 0.676. The median value and mean value of big firms are 0.127 and 0.178, respectively. Mean value > median

value indicates that the data has asymmetric distribution, the data is likely to appear outliers. The standard deviation for big firms is 0.158, indicating a variation in cash holdings of 0.158. This indicator is smaller than the mean value, indicating that the volatility is not high.

- Small: Min value = 0.00103 shows that in 50 small firms, the lowest cash holding level is 0.00103. Max value = 0.601, indicating that the highest cash holdings in small firms is 0.601. Median value and mean value of these firms are 0.0592 and 0.125, respectively. Mean value > median value shows that the data has asymmetric distribution. The difference between these two values is 0.0658, which is likely to appear outliers. The standard deviation for small firms is 0.150, indicating a variation in cash holdings of 0.160. A standard deviation larger than the mean value indicates a large degree of variation.

3.4.2. Groups of above/below median of the continuous variable

```
df %>%
  mutate(median_lev = ifelse(leverage > median(leverage),
                             "above median of leverage",
                             "below median of leverage")) %>%
  group_by(median_lev) %>%
  summarize(min = min(cash_holding),
            max = max(cash_holding),
            median = median(cash_holding),
            mean = mean(cash_holding),
            standard_deviation = sd(cash_holding))
```

median_lev	min	max	median	mean	standard_deviation
<chr>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
above median of leverage	0.000650	0.524	0.0833	0.118	0.126
below median of leverage	0.00157	0.676	0.126	0.183	0.174

- Firms with a financial leverage ratio higher than the median value : Min value = 0.000650, the lowest cash holding ratio of firms with a financial leverage ratio higher than the median is 0.000650. Max value = 0.524, the highest cash holding ratio in these firms is 0.524. Median value and mean value of these firms are 0.833 and 0.118, respectively. Mean value > median value indicates that the data has asymmetric distribution, the data is likely to appear outliers. The standard deviation for high leverage firms is 0.126, which shows a variation in cash holdings of 0.126.
- Firms with financial leverage ratio lower than median value : Min value = 0.00157, the lowest cash holding ratio of firms with financial leverage ratio higher than median is 0.00157. Max value = 0.676, the highest cash holding ratio in these firms is 0.676. Median value and mean value of these firms are 0.126 and 0.183, respectively. Mean value > median value indicates that the data has asymmetric

distribution, the data is likely to appear outliers. The standard deviation for high leverage firms is 0.174, which shows a variation in cash holdings of 0.174.

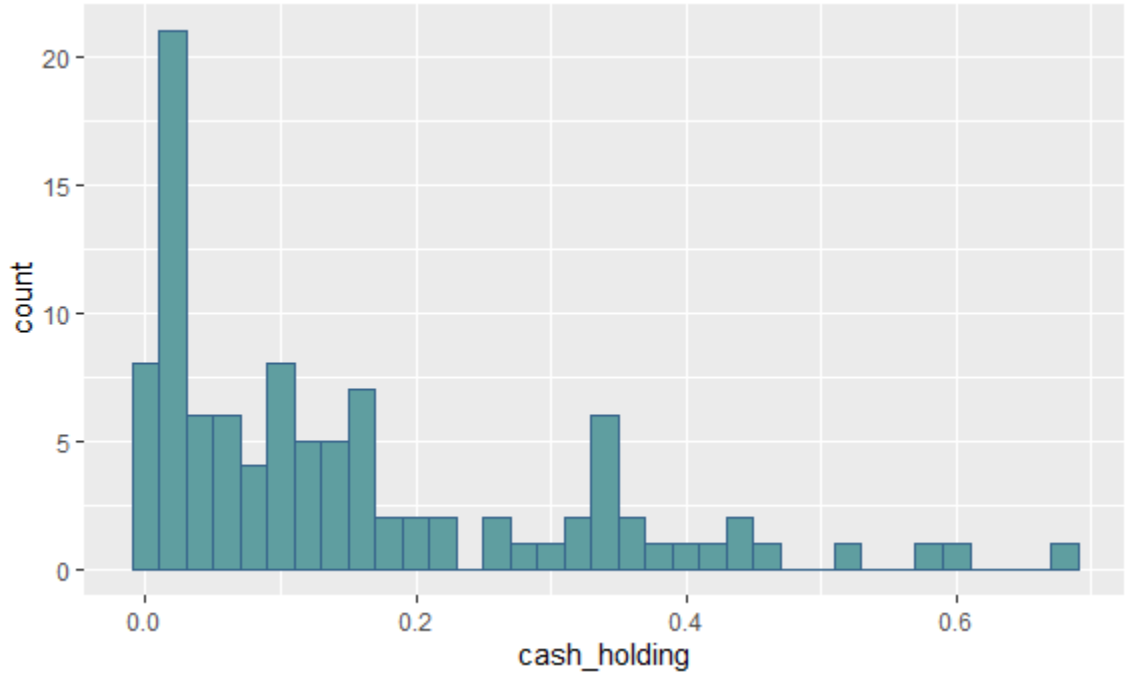
- The above median of leverage group has both mean and median of cash holdings lower than the below median of leverage group. The min and max values of the above median of leverage group are also lower than those of the below median of leverage group. This shows that, high leverage will have low cash holding. So, there is a negative correlation between leverage and cash holding.

4. Data visualization

4.1. Histogram plot of cash holding

Use ggplot2 + geom_histogram function to draw histogram plot, change the binwidth, fill, color arguments to set the width of bin to 0.2, the color of the column and the color of the column border.

```
ggplot(df, aes(x=cash_holding)) +
  geom_histogram(binwidth=0.02, fill='cadetblue', color='steelblue4')
```



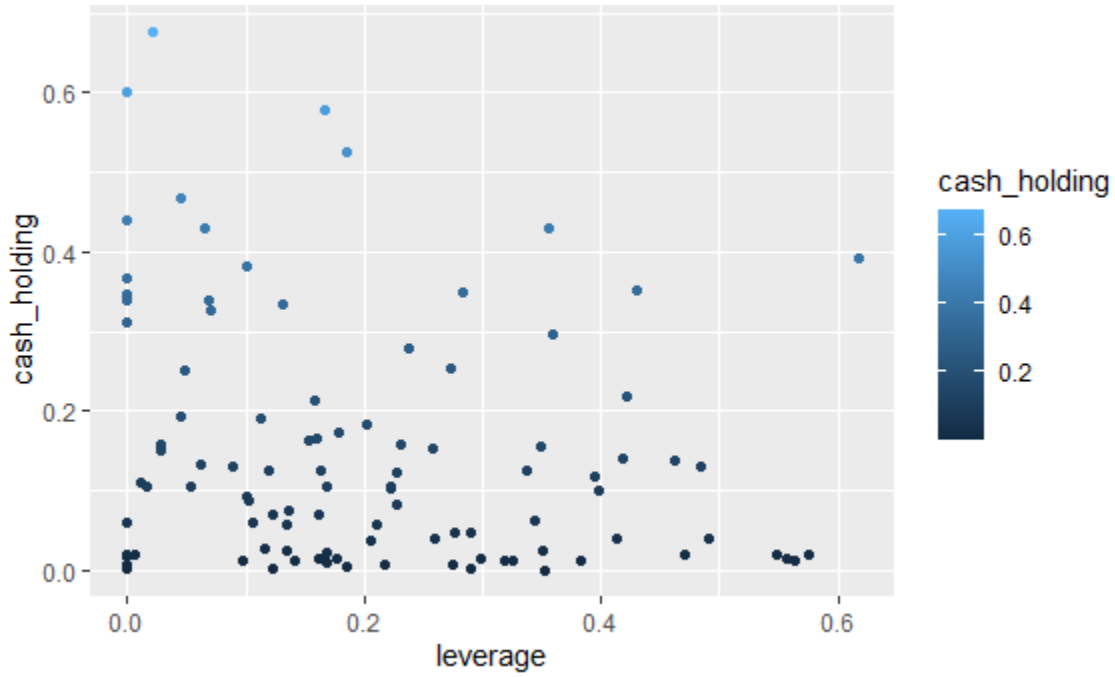
According to the above plot, most firms with cash holdings of 0.2 or less, cash holding ratio from 0.2 to 0.4 have fewer firms, very few firms have cash holdings from 0.4-0.6, and only about 1-2 The company has a cash holding above 0.6. The 3 cash holding

ratios with the most firms owned are 0.01-0.03 with 21 firms, 0.9-0.11 with 8 firms, and 0.15-0.17 with 7 firms. It can be seen that firms tend to hold cash at a low level.

4.2. Scatter plot of cash holding with the continuous variable

To plot the scatter, use ggplot +geom_point, with the horizontal axis being the continuous variable - leverage, the vertical axis being the cash holding.

```
ggplot(df, aes(x= leverage, y= cash_holding, color = cash_holding))
+geom_point()
```

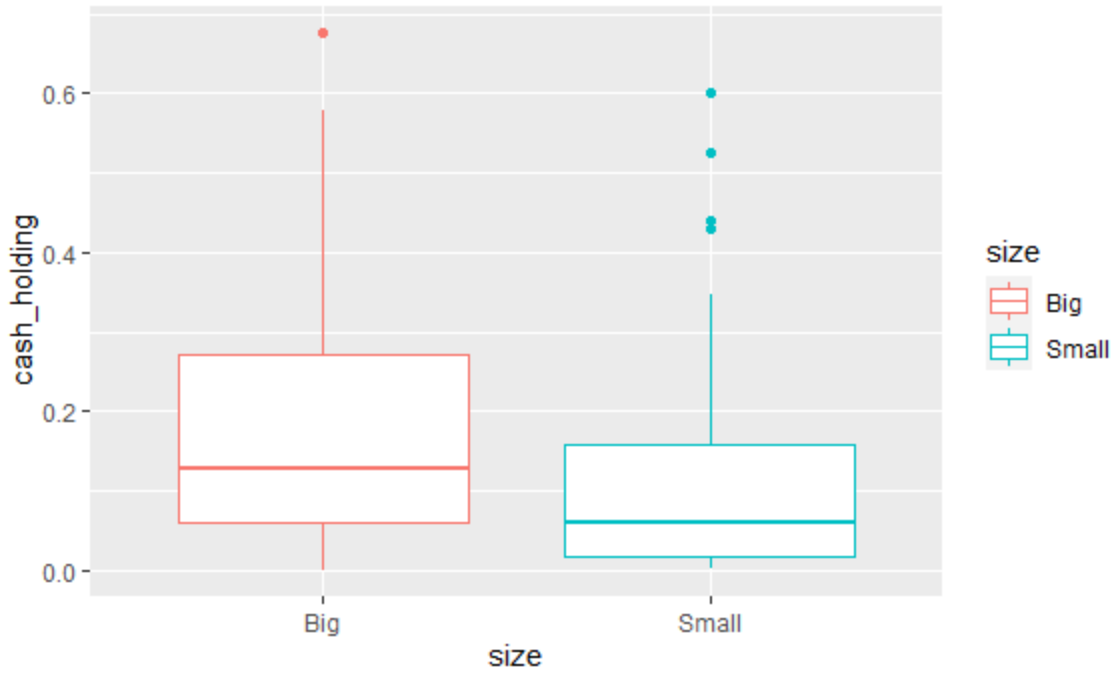


The above above shows that the higher the leverage, the lower the cash holding. It can be seen that leverage is negatively correlated with cash holding.

4.3. Boxplot of cash holding with the discrete variable

Use the ggplot+geom_plot function to draw a boxplot chart, with the horizontal axis being the discrete variable – size, the vertical axis being the cash holding. The color of the chart is displayed according to the size variable.

```
df %>% ggplot(aes(x=size, y = cash_holding, color = size)) +geom_boxplot()
```

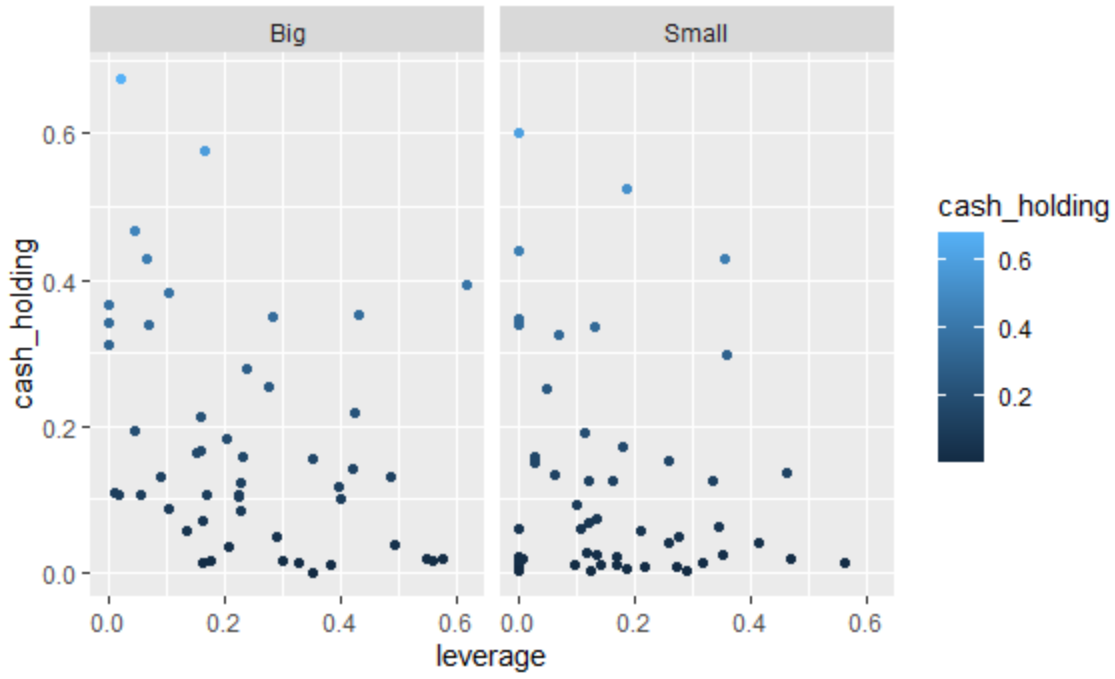


The data in both big and small tend to skew below the mean. The distance between min and max value big is larger than that of small, indicating that big's cash holding is more dispersed than small's. The interquartile range (Q3-Q1) of big is larger than that of big, showing that big's cash holding has a smaller variability than small. In both big and small there are outliers, however small has more outliers than big. Both the median and mean values of big are larger than those of small, which means that large firms will hold more cash than small firms. Therefore, firm size and cash holding have a positive correlation.

4.4. Plot that allow the combination of continuous, discrete variables and cash holding

Use the facet_wrap() function to draw a graph combining 3 variables, with the horizontal axis being a continuous variable – leverage, the vertical axis being cash holding. The chart is divided by the independent variable – size.

```
ggplot(df, aes(x = leverage, y = cash_holding, color=cash_holding)) +
  geom_point() +
  facet_wrap(~ size, nrow = 1)
```



As shown in the above plot, both small and big show that leverage increases will make cash holding decrease. This shows that, in firms of different sizes, there is a negative correlation between leverage and cash holding.

In small, there are more firms with cash holding concentration below 0.1 than in big (it can be seen that small firms tend to hold less cash). In big, there are more firms with high cash holdings (over 0.4) than in small (showing that large-scale firms hold more cash). Therefore, the larger the company size, the higher the cash holding. There is a positive correlation between firm size and cash holding.

5. Using loop

5.1. Count the number of firms in an industry

First, use the group_by and the summarise function to create a new tibble containing the names of the industries and the number of firms in these industries.

```
df2= df %>%
  group_by(industry) %>%
  summarise(n = n())
```

```
   industry                 n
   <chr>                <int>
 1 Basic Materials          9
 2 Consumer Cyclicals      15
 3 Consumer Non-Cyclicals   9
 4 Energy                   7
 5 Financials               7
 6 Healthcare               5
 7 Industrials             27
 8 Real Estate             13
 9 Technology               1
10 Utilities                7
```

Next is the command asking the user to enter the business name he wants.

```
industry_name = readline("Please enter an industry name: ")
```

Then create a loop using the repeat function to output the company number of the industry entered.

```
repeat{
  if (industry_name %in% df$industry){
    print(paste(x,   "has",   df2[which(df2$industry==industry_name),   'n'],
"firm"))
    y = readline("Do you want to continue? (y/n): ")
    if (y=="y"){
      new_industry_name = readline("Please enter an new industry name: ")
    } else if (y=="n"){
      print("Thank you! The loop ends.")
      break}
  } else {
    industry_name = readline("The industry name is incorrect. Please enter
again
                  (or if you don't want to continue, please enter 'n' to end:
")
    if (industry_name =="n"){
      break
    } else {next}
  }
  x = new_industry_name
}
```

Create a statement that checks if the user entered the correct industry name. (`if (industry_name %in% df$industry)`).

```
> industry_name = readline("Please enter an industry name: ")
Please enter an industry name: Industrials
```

- If the industry name is entered correctly:
  The program will print to the screen: the name of the industry and the number of firms in that industry. Then ask if the user wants to continue the loop

```
[1] "Industrials has 27 firm"
Do you want to continue? (y/n):
```

If the user presses "y", the program will ask for the trade name and continue the loop until the user presses "n" to end.

```
[1] "Industrials has 27 firm"
Do you want to continue? (y/n): y
Please enter an new industry name: Utilities
[1] "Utilities has 27 firm"
Do you want to continue? (y/n):
```

If the user presses “n” the program will end the loop:

```
Do you want to continue? (y/n): n
[1] "Thank you! The loop ends."
```

If the industry name is entered incorrectly: the program will print the error message on the screen and ask to re-enter:

```
> industry_name = readline("Please enter an industry name: ")
Please enter an industry name: n
The industry name is incorrect. Please enter again
                  (or if you don't want to continue, please enter 'n' to end:
```

If the user re-enters the correct industry name, the program will print to the screen the industry name and company number, and then continue the loop as the process when the industry name is correct.

```
The industry name is incorrect. Please enter again
                 (or if you don't want to continue, please enter 'n' to end: Financials
[1] "Financials has 7 firm"
Do you want to continue? (y/n):
```

If the re-entered industry name is incorrect, the program will make a re-entry request until it is correct,

```
The industry name is incorrect. Please enter again
                 (or if you don't want to continue, please enter 'n' to end: indus
The industry name is incorrect. Please enter again
                 (or if you don't want to continue, please enter 'n' to end: finan
The industry name is incorrect. Please enter again
                 (or if you don't want to continue, please enter 'n' to end:
```

or the user can press “n” to end the loop:

```
The industry name is incorrect. Please enter again
                  (or if you don't want to continue, please enter 'n' to end: n
[1] "Thank you! The loop ends."
```

5.2. Count the number of firms in an industry and with cash holding above a certain value

First, create a statement that requires the user to enter a value. (example: value = 0.5)

```
value = (readline("Please enter a value: "))
```

```
Please enter a value: 0.5
```

```
df3 = df %>%
  group_by(industry) %>%
  filter(cash_holding>value) %>%
  summarise(n = n())
```

The above command will create a new tibble with the value entered:

	industry	n
	<chr>	<int>
1	Consumer Cyclicals	1
2	Consumer Non-Cyclicals	2
3	Healthcare	1

Then the statement asks the user to enter the industry name and executes the loop:

```
industry_name_2 = readline("Please enter an industry name: ")
repeat{
  if (industry_name_2 %in% df3$industry){
    count = df3[which(df3$industry==industry_name_2), 'n']
    print(paste(industry_name_2, "with cash holding are",  value,"has", count,
"firm"))
    y = readline("Do you want to continue? (y/n): ")
    if (y=="y"){
      new_industry_name = readline("Please enter an new industry name: ")
    } else if (y=="n"){
      print("Thank you! The loop ends.")
      break}
  } else {
    industry_name_2 = readline("The industry name is incorrect. Please enter
again
                  (or if you don't want to continue, please enter 'n' to end:
")
    if (industry_name_2 =="n"){
      break
    } else {next}
  }
  industry_name_2 = new_industry_name
}
```

- If the industry name is entered correctly:
  - The industry name is not in the list of firms with cash holdings greater than the input value: the program will print a screen stating that there are no firms in that industry.

    ```
    [1] "No firms in the Financials industry has a cash holding above  0.5"
    ```

  - The industry name is in the list of firms with cash holding larger than the input value, the program will print to the screen: industry name, value and company number:

    ```
    "Healthcare with cash holding are 0.5 has 1 firm"
    ```

  - Then the program will continue the same loop as 5.1.
- If the industry name is entered incorrectly:
  The program will perform the same procedures as 5.1.

REFERENCE

Chen, S., & Liu, S. (2013). “Corporate cash holdings: Study of Chinese firms”. Retrieved 10 May 2022, from https://summit.sfu.ca

Kafayat, A., Rehman, K. U., & Farooq, M. (2014). Factors effecting corporate cash holding of non-financial firms in Pakistan. *Acta Universitatis Danubius. Œconomica*, *10*(3).

Ozkan, A., & Ozkan, N. (2004). Corporate cash holdings: An empirical investigation of UK firms. Journal of banking & finance, 28(9), 2103-2134.

Saddour, K. (2006). *The determinants and the value of cash holdings: Evidence from French firms* (No. halshs-00151916).