AM16_Financial_Reporting_Analytics_Assignment_2_Group_

Group 7

07/03/2020

Contents

Part A 1
Question 1 1
Question 2 10
Question 3 11
Question 4 12
EBITDA margin Comparison 12
Quick ratio Comparison 13
EV_to_EBITDA Comparison 14
Question 5 15

Part B 15

Part A

Question 1

Read in data

```r
# Load in airline Data
airline_data_raw <- read.delim("assignment_2_files/airline_data.txt", header = TRUE)
```

```r
# Overview of data
glimpse(airline_data_raw)
```

```
## Observations: 738
## Variables: 51
## $ gvkey    <int> 1045, 1045, 1045, 1045, 1045, 1045, 1045, 1045, 1045,...
## $ datadate <int> 20001231, 20011231, 20021231, 20031231, 20041231, 200...
## $ fyear    <int> 2000, 2001, 2002, 2003, 2004, 2005, 2006, 2007, 2008,...
## $ indfmt   <fct> INDL, INDL, INDL, INDL, INDL, INDL, INDL, INDL, INDL,...
## $ consol   <fct> C, C, C, C, C, C, C, C, C, C, C, C, C, C, C, C, C, C,...
## $ popsrc   <fct> D, D, D, D, D, D, D, D, D, D, D, D, D, D, D, D, D, D,...
## $ datafmt  <fct> STD, STD, STD, STD, STD, STD, STD, STD, STD, STD, STD...
## $ tic      <fct> AAL, AAL, AAL, AAL, AAL, AAL, AAL, AAL, AAL, AAL, AAL...
## $ conm     <fct> AMERICAN AIRLINES GROUP INC, AMERICAN AIRLINES GROUP ...
## $ curcd    <fct> USD, USD, USD, USD, USD, USD, USD, USD, USD, USD, USD...
## $ aco      <dbl> 886.0, 1312.0, 96.0, 237.0, 240.0, 334.0, 225.0, 638....
## $ act      <dbl> 5179.0, 6540.0, 4937.0, 4682.0, 4971.0, 6164.0, 6902....
## $ aqc      <dbl> NA, 742, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, -206, 0, 0,...
## $ capx     <dbl> 3678.0, 3640.0, 1881.0, 680.0, 1027.0, 681.0, 530.0, ...
## $ ceq      <dbl> 7176.0, 5373.0, 957.0, 46.0, -581.0, -1478.0, -606.0,...
## $ che      <dbl> 2233.0, 2992.0, 2733.0, 3133.0, 3407.0, 4324.0, 5183....
## $ cogs     <dbl> 17120.0, 19419.0, 18555.0, 14430.0, 15120.0, 16832.0,...
## $ cshfd    <dbl> 162.000, 154.000, 156.000, 158.000, 161.000, 165.000,...
## $ csho     <dbl> 152.063, 154.485, 156.089, 159.582, 161.156, 182.732,...
## $ cshpri   <dbl> 150.000, 154.000, 156.000, 158.000, 161.000, 165.000,...
## $ dlc      <dbl> 796.0, 772.0, 868.0, 804.0, 806.0, 1239.0, 1349.0, 10...
## $ dltt     <dbl> 5474.0, 9834.0, 12310.0, 13126.0, 13524.0, 13456.0, 1...
## $ dp       <dbl> 1202.0, 1404.0, 1366.0, 1377.0, 1292.0, 1164.0, 1157....
## $ epsfi    <dbl> 5.03, -11.43, -22.57, -7.76, -4.74, -5.21, 0.98, 1.78...
## $ epsfx    <dbl> 4.81, -11.43, -16.22, -7.76, -4.74, -5.21, 0.98, 1.78...
## $ gdwlip   <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ ib       <dbl> 779.0, -1762.0, -2523.0, -1228.0, -761.0, -861.0, 231...
## $ ibadj    <dbl> 779.0, -1762.0, -2523.0, -1228.0, -761.0, -861.0, 231...
## $ invt     <dbl> 757.0, 822.0, 627.0, 516.0, 488.0, 515.0, 506.0, 601....
## $ lct      <dbl> 6990.0, 7512.0, 7240.0, 6559.0, 7018.0, 8320.0, 8505....
## $ mib      <dbl> NA, 0, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA...
## $ niadj    <dbl> 813.0, -1762.0, -3511.0, -1228.0, -761.0, -861.0, 231...
## $ oiadp    <dbl> 1381.0, -1860.0, -2622.0, -795.0, -133.0, 102.0, 1060...
## $ oibdp    <dbl> 2583.0, -456.0, -1256.0, 582.0, 1159.0, 1266.0, 2217....
## $ pstk     <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,...
## $ recd     <dbl> 27.0, 52.0, 66.0, 62.0, 59.0, 60.0, 45.0, 41.0, 49.0,...
## $ rect     <dbl> 1303.0, 1414.0, 1481.0, 796.0, 836.0, 991.0, 988.0, 1...
## $ rectr    <dbl> 1303.0, 1414.0, 858.0, 796.0, 836.0, 991.0, 988.0, 10...
## $ sale     <dbl> 19703.0, 18963.0, 17299.0, 17440.0, 18645.0, 20712.0,...
## $ seq      <dbl> 7176.0, 5373.0, 957.0, 46.0, -581.0, -1478.0, -606.0,...
## $ txt      <dbl> 508.0, -994.0, -1337.0, -80.0, 0.0, 0.0, 0.0, 0.0, 0....
## $ wdp      <dbl> NA, NA, NA, NA, NA, NA, NA, NA, -1122.0, -88.0, -28.0...
## $ xint     <dbl> 467.0, 538.0, 685.0, 703.0, 871.0, 957.0, 1030.0, 914...
## $ xpp      <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, NA, NA, NA, NA...
## $ xsga     <dbl> NA, NA, NA, 2428.0, 2366.0, 2614.0, 2687.0, 2777.0, 3...
## $ exchg    <int> 14, 14, 14, 14, 14, 14, 14, 14, 14, 14, 14, 14, 14, 1...
## $ cik      <int> 6201, 6201, 6201, 6201, 6201, 6201, 6201, 6201, 6201,...
## $ costat   <fct> A, A, A, A, A, A, A, A, A, A, A, A, A, A, A, A, A, A,...
## $ prcc_f   <dbl> 39.1875, 22.3000, 6.6000, 12.9500, 10.9500, 22.2300, ...
## $ loc      <fct> USA, USA, USA, USA, USA, USA, USA, USA, USA, USA, USA...
## $ sic      <int> 4512, 4512, 4512, 4512, 4512, 4512, 4512, 4512, 4512,...
```

```
# Definition of variables in use

# gvkey = Global Company Key - Company Annual Descriptor
# datadate = Data Date - Company Annual Descriptor
# fyear = Fiscal Year
# indfmt = Industry Format
# consol = Level of Consolidation - Company Annual Descriptor
# popsrc = Population Source
# datafmt = Data Format - Company Annual Descriptor
# tic = Ticker
# conm = Company Name
# curcd = Currency
# aco = Current Assets - Other - Total
# act = Current Assets - Total
# aqc = Acquisitions
# capx = Capital Expenditures
# ceq = Common/Ordinary Equity - Total
# che = Cash and Short-Term Investments- Total
# cogs = Cost of Goods Sold
# cshfd = Common Shares Used to Calculate EPS Diluted
# csho = Common Shares Outstanding
# cshpri = Common Shares Used to Calculate EPS Basic
# dlc = Debt in Current Liabilities
# dltt = Long-Term Debt - Total
# dp = Depreciation and Amortization - Total
# epsfi = Earnings per Share - Diluted  Including Extraordinary Items
# epsfx = Earnings per Share - Diluted  Excluding Extraordinary Items
# gdwlip = Impairment of Goodwill Pretax
# ib = Income Before Extraordinary Items
# ibadj = Income Before Extraordinary Items - Adjusted for Common Stock Equivalents
# invt = Inventories - Total
# lct = Current Liabilities - Total
# mib = Minority Interest - Balance Sheet
# niadj = Net Income (Loss)
# oiadp = Operating Income After Depreciation
# oibdp = Operating Income Before Depreciation
# pstk = Preferred/Preference Stock (Capital) - Total
# recd = Receivables-Est Doubtful
# rect = Receivables - Total
# rectr = Receivables - Trade
# sale = Sales/Turnover (Net)
# seq = Stockholders' Equity - Total
# txt = Income Taxes - Total
# wdp = Writedowns Pretax
# xint = Interest and Related Expense
# xpp = Prepaid Expenses and Accrued Income
# xsga = Selling, General and Administrative Expenses
# exchg =
# cik  =
# costat =
# prcc_f = Share price
# loc =
# sic =
```

Clean Data

```
# Cleaning data
airline_data <- airline_data_raw %>%
  rename(date =datadate) %>%
  mutate(date = ymd(date))
```

Brief Summary of Data

```
# Summary statistics
summary(airline_data)
```

```
##      gvkey             date                 fyear       indfmt    consol
##  Min.   :  1045   Min.   :2000-03-31   Min.   :1999   INDL:738   C:738
##  1st Qu.: 12392   1st Qu.:2004-12-31   1st Qu.:2004
##  Median : 63026   Median :2008-12-31   Median :2008
##  Mean   : 84099   Mean   :2009-08-05   Mean   :2009
##  3rd Qu.:152011   3rd Qu.:2013-12-31   3rd Qu.:2013
##  Max.   :298636   Max.   :2019-12-31   Max.   :2019
##
```

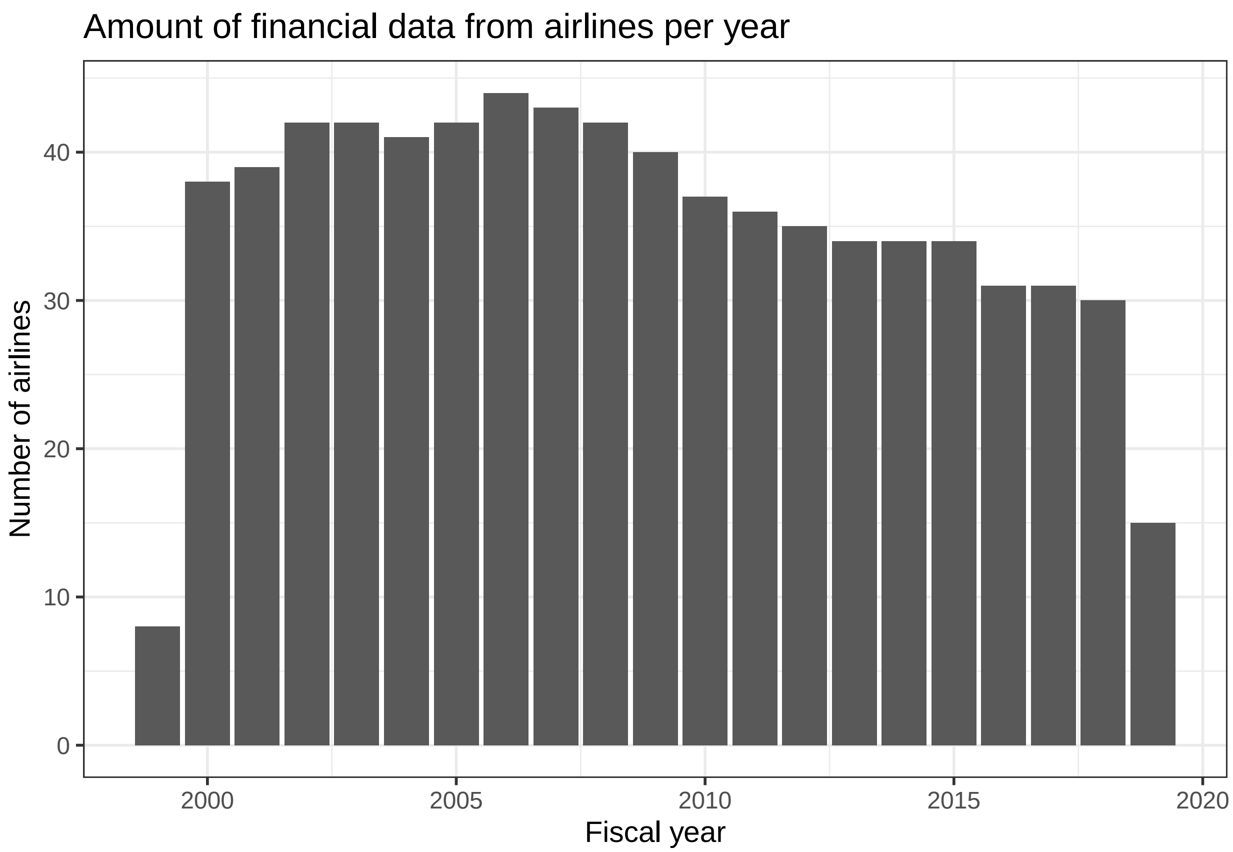
```
##  popsrc  datafmt        tic                              conm
##  D:738   STD:738   ICAGY  : 21   INTL CONSOL AIRLINES GROUP : 21
##                    SKYW   : 21   SKYWEST INC                : 21
##                    AAL    : 20   ALASKA AIR GROUP INC       : 20
##                    ALK    : 20   AMERICAN AIRLINES GROUP INC: 20
##                    DAL    : 20   DELTA AIR LINES INC        : 20
##                    HA     : 20   HAWAIIAN HOLDINGS INC      : 20
##                    (Other):616   (Other)                    :616
##  curcd          aco              act              aqc
##  CAD: 67   Min.   :   0.00   Min.   :    0.0   Min.   :-3698.00
##  USD:671   1st Qu.:  25.92   1st Qu.:  273.3   1st Qu.:    0.00
##            Median : 126.80   Median : 1124.2   Median :    0.00
##            Mean   : 323.33   Mean   : 2640.4   Mean   :    0.77
##            3rd Qu.: 481.82   3rd Qu.: 4127.0   3rd Qu.:    0.00
##            Max.   :5086.00   Max.   :19257.0   Max.   : 1951.00
##            NA's   :16        NA's   :17        NA's   :42
##       capx              ceq                 che
##  Min.   :  -6.998   Min.   :-25869.00   Min.   :    0.0
##  1st Qu.:  76.716   1st Qu.:    81.27   1st Qu.:  149.8
##  Median : 348.553   Median :   577.51   Median :  663.6
##  Mean   : 828.572   Mean   :  1448.38   Mean   : 1502.2
##  3rd Qu.:1084.544   3rd Qu.:  2144.55   3rd Qu.: 2362.0
##  Max.   :6151.000   Max.   : 16652.15   Max.   :10286.0
##  NA's   :18         NA's   :27          NA's   :16
##       cogs            cshfd              csho             cshpri
##  Min.   :    0.0   Min.   :   0.01   Min.   :   0.00   Min.   :   0.01
##  1st Qu.:  785.4   1st Qu.:  42.41   1st Qu.:  31.32   1st Qu.:  42.05
##  Median : 2297.0   Median :  87.48   Median :  71.34   Median :  86.15
##  Mean   : 7052.0   Mean   : 195.43   Mean   : 188.87   Mean   : 193.84
##  3rd Qu.:11901.9   3rd Qu.: 258.78   3rd Qu.: 241.24   3rd Qu.: 252.60
##  Max.   :36927.7   Max.   :6479.06   Max.   :7555.72   Max.   :6479.06
##  NA's   :17        NA's   :86        NA's   :28        NA's   :68
##       dlc              dltt               dp              epsfi
##  Min.   :   0.00   Min.   :    0.0   Min.   :   0.00   Min.   :-182.2900
##  1st Qu.:  36.83   1st Qu.:  242.7   1st Qu.:  39.35   1st Qu.:  -0.2800
##  Median : 189.59   Median : 1369.6   Median : 191.92   Median :   0.7100
##  Mean   : 623.56   Mean   : 3363.3   Mean   : 515.23   Mean   :  -0.2609
##  3rd Qu.: 827.68   3rd Qu.: 5427.6   3rd Qu.: 903.00   3rd Qu.:   2.1000
##  Max.   :7428.34   Max.   :29081.0   Max.   :3183.96   Max.   : 178.5700
##  NA's   :16        NA's   :16        NA's   :17        NA's   :68
##      epsfx              gdwlip                ib
##  Min.   :-182.2900   Min.   :-7253.000   Min.   :-21176.00
##  1st Qu.:  -0.2800   1st Qu.:  -85.523   1st Qu.:   -15.32
##  Median :   0.7450   Median :   -8.917   Median :    51.30
##  Mean   :  -0.2157   Mean   : -343.407   Mean   :   146.30
##  3rd Qu.:   2.0700   3rd Qu.:   -2.773   3rd Qu.:   281.00
##  Max.   : 178.5700   Max.   :   64.000   Max.   : 22876.00
##  NA's   :68          NA's   :690         NA's   :17
##      ibadj                invt              lct                mib
##  Min.   :-21186.00   Min.   :   0.00   Min.   :    0.0   Min.   :  0.00
##  1st Qu.:   -15.40   1st Qu.:  15.60   1st Qu.:  227.6   1st Qu.:  0.00
##  Median :    51.16   Median :  53.16   Median : 1112.0   Median :  0.00
##  Mean   :   145.51   Mean   : 190.48   Mean   : 3396.5   Mean   : 14.93
##  3rd Qu.:   281.00   3rd Qu.: 241.67   3rd Qu.: 5554.7   3rd Qu.:  0.00
##  Max.   : 22866.00   Max.   :1851.00   Max.   :20204.0   Max.   :413.65
##  NA's   :17          NA's   :27        NA's   :16        NA's   :32
##      niadj               oiadp               oibdp
##  Min.   :-21186.00   Min.   :-2945.000   Min.   :-1997.00
##  1st Qu.:   -17.00   1st Qu.:    3.363   1st Qu.:   85.94
##  Median :    48.95   Median :  132.304   Median :  344.00
##  Mean   :   141.33   Mean   :  447.962   Mean   :  963.19
##  3rd Qu.:   281.00   3rd Qu.:  491.580   3rd Qu.: 1210.00
##  Max.   : 22866.00   Max.   : 7837.000   Max.   : 9672.00
##  NA's   :17          NA's   :17          NA's   :17
##       pstk              recd              rect             rectr
##  Min.   :   0.00   Min.   :  0.000   Min.   :    0.0   Min.   :   0.00
##  1st Qu.:   0.00   1st Qu.:  0.525   1st Qu.:   33.4   1st Qu.:  28.79
##  Median :   0.00   Median :  4.200   Median :  133.1   Median : 101.00
##  Mean   :  39.64   Mean   : 17.917   Mean   :  623.6   Mean   : 502.34
##  3rd Qu.:   0.00   3rd Qu.: 22.758   3rd Qu.:  807.5   3rd Qu.: 753.89
##  Max.   :3988.00   Max.   :822.000   Max.   :12421.0   Max.   :4861.24
##  NA's   :27        NA's   :209       NA's   :16        NA's   :59
##       sale              seq                 txt
##  Min.   :    0.0   Min.   :-25869.00   Min.   :-8013.000
##  1st Qu.:  998.4   1st Qu.:    91.51   1st Qu.:   -0.062
##  Median : 3034.0   Median :   585.82   Median :   17.156
##  Mean   : 8832.0   Mean   :  1477.49   Mean   :   42.103
##  3rd Qu.:13916.0   3rd Qu.:  2132.25   3rd Qu.:   74.100
##  Max.   :47007.0   Max.   : 16652.15   Max.   : 2631.000
##  NA's   :17        NA's   :16          NA's   :17
##       wdp                 xint              xpp
##  Min.   :-1142.000   Min.   :   0.00   Min.   :   0.000
##  1st Qu.:  -89.344   1st Qu.:  20.66   1st Qu.:   5.057
##  Median :  -21.483   Median :  89.95   Median :  33.505
##  Mean   :  -99.911   Mean   : 214.61   Mean   :  94.060
##  3rd Qu.:   -5.094   3rd Qu.: 336.88   3rd Qu.:  99.955
##  Max.   :  127.000   Max.   :1278.00   Max.   :1167.070
##  NA's   :581         NA's   :18        NA's   :422
##       xsga             exchg            cik          costat
##  Min.   :   0.00   Min.   : 0.00   Min.   :   3202   A:531
##  1st Qu.:  99.16   1st Qu.:11.00   1st Qu.: 766421   I:207
##  Median : 374.77   Median :14.00   Median :1030475
##  Mean   :1094.64   Mean   :13.36   Mean   : 863328
##  3rd Qu.:1499.63   3rd Qu.:19.00   3rd Qu.:1166291
##  Max.   :8938.00   Max.   :19.00   Max.   :1614436
##  NA's   :200                       NA's   :62
##      prcc_f              loc           sic
##  Min.   :  0.0014   USA    :457   Min.   :4512
##  1st Qu.:  6.6200   CAN    : 67   1st Qu.:4512
##  Median : 14.4800   CHN    : 38   Median :4512
##  Mean   : 21.7204   BRA    : 29   Mean   :4512
##  3rd Qu.: 27.5800   ESP    : 21   3rd Qu.:4512
##  Max.   :167.8300   IRL    : 20   Max.   :4512
##  NA's   :95         (Other):106
```

```r
# Number of observations of financial data for a given airline
per_airline <- airline_data %>%
  group_by(tic,conm) %>%
  summarise(counted = n())
print(per_airline)
```

```
## # A tibble: 58 x 3
## # Groups:   tic [58]
##    tic    conm                          counted
##    <fct>  <fct>                           <int>
##  1 3UAIRQ US AIRWAYS GROUP INC-OLD            5
##  2 6039B  GLOBAL AVIATION HOLDINGS INC       11
##  3 AAI    AIRTRAN HOLDINGS INC               11
##  4 AAL    AMERICAN AIRLINES GROUP INC        20
##  5 AFLYY  AIR FRANCE - KLM                   17
##  6 AIDEF  AIR CANADA                         15
##  7 ALGT   ALLEGIANT TRAVEL CO                14
##  8 ALK    ALASKA AIR GROUP INC               20
##  9 ALK1   ALASKA AIRLINES INC                10
## 10 AMR1   AMERICAN AIRLINES INC              19
## # ... with 48 more rows
```

```
rm(per_airline)
```

```
# Amount of financial data per year
per_year <- airline_data %>%
  group_by(fyear) %>%
  summarise(counted = n())

ggplot(per_year,aes(x=fyear, y=counted))+
  geom_col()+
  theme_bw()+
  labs(title="Amount of financial data from airlines per year",
       x="Fiscal year",
       y="Number of airlines")+
  theme(
    legend.position= "none"
  )
```



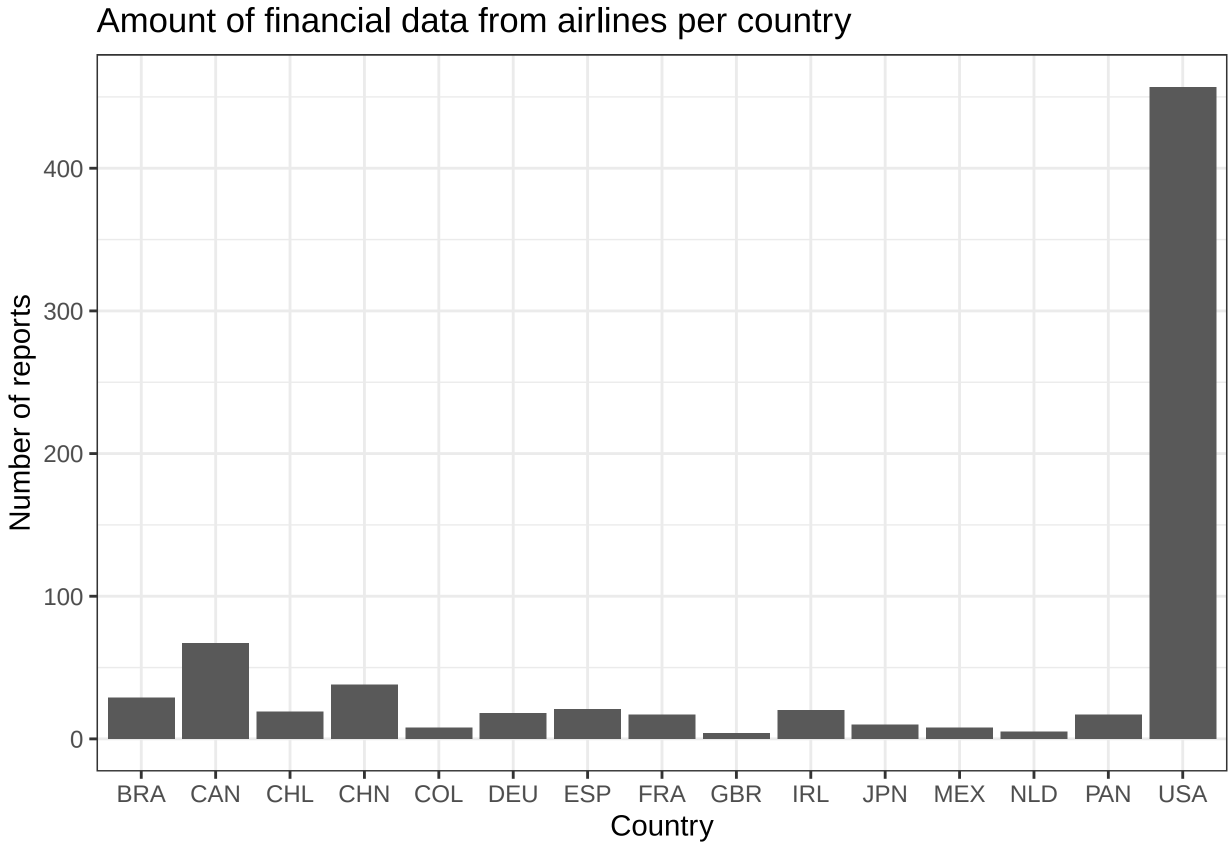
```r
rm(per_year)
```

```r
# Different currencies
currency <- airline_data %>%
  group_by(curcd) %>%
  summarise(counted = n())
print(currency)
```

```
## # A tibble: 2 x 2
##   curcd counted
##   <fct>   <int>
## 1 CAD        67
## 2 USD       671
```

```r
rm(currency)
```

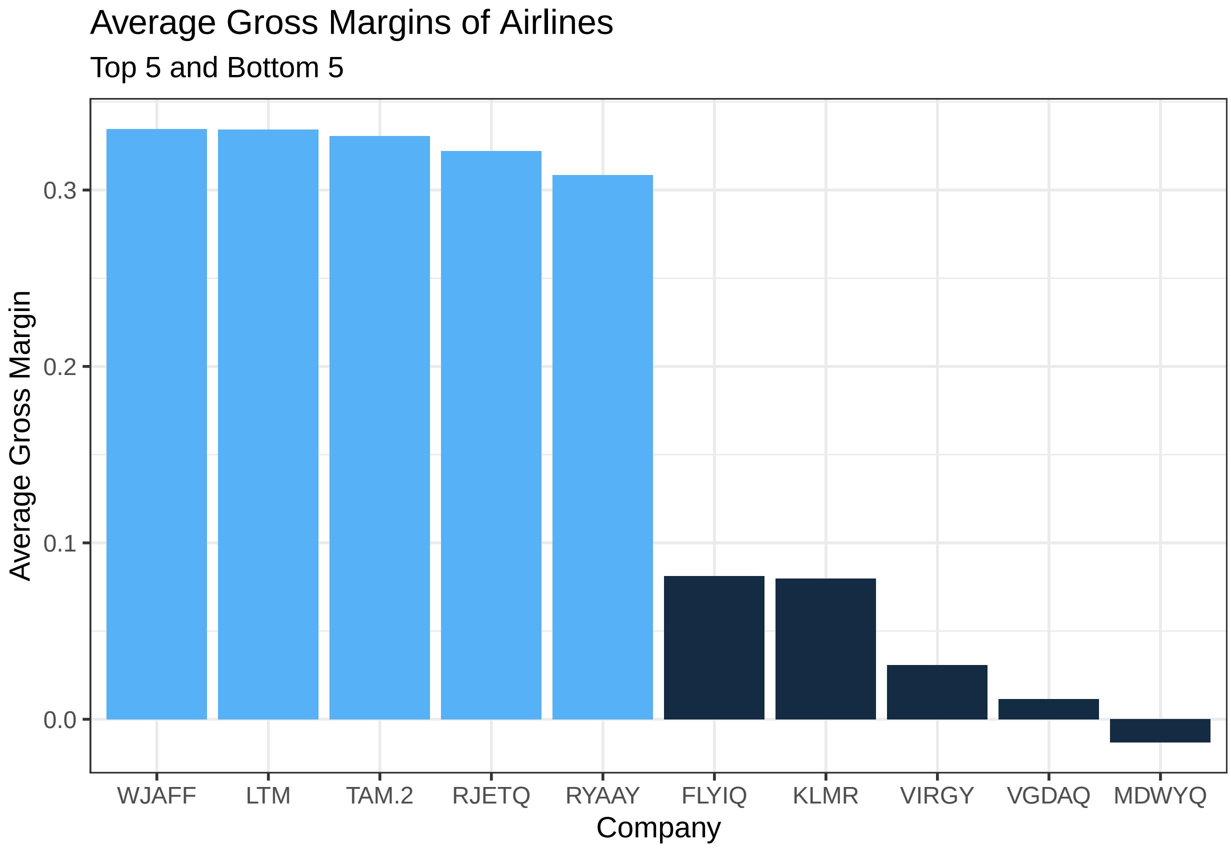
```r
# Amount of financial data by country
location <- airline_data %>%
  group_by(loc) %>%
  summarise(counted = n())

ggplot(location,aes(x=loc, y=counted))+
  geom_col()+
  theme_bw()+
  labs(title="Amount of financial data from airlines per country",
       x="Country",
       y="Number of reports")+
  theme(
    legend.position= "none"
  )
```



```
rm(location)
```

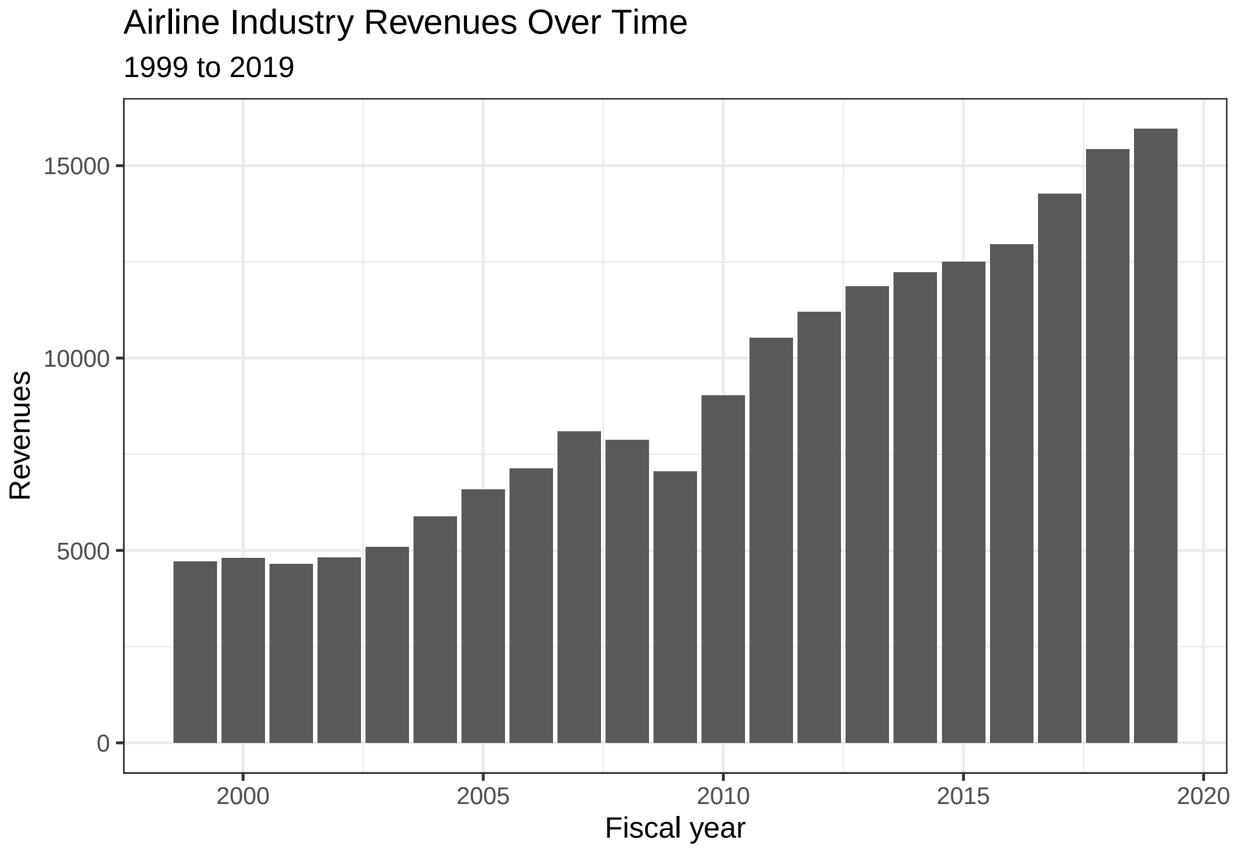
```
# Visualizing the aveage gross margin - top and bottom 5

# Calculate gross margin and average gross margin
gross_margin <- airline_data %>%
  mutate(gross_margin = (sale-cogs)/sale) %>%
  group_by(tic) %>%
  summarise(average_gross_margin = mean(gross_margin))

# Select top 5
top_gross_margin <- gross_margin %>%
  top_n(5, average_gross_margin) %>%
  mutate(top=1)

# Select bottom 5
bottom_gross_margin <- gross_margin %>%
  filter(average_gross_margin > -999) %>%
  top_n(5, -average_gross_margin) %>%
  mutate(top = 0)

# Combine the two
```

```r
top_bottom_gross_margrin <- rbind(top_gross_margin,bottom_gross_margin)

# Plot
ggplot(top_bottom_gross_margrin,aes(x=reorder(tic, -average_gross_margin), y=average_gross_margin))+
  geom_col(aes(fill=top))+
  theme_bw()+
  labs(title="Average Gross Margins of Airlines",
       subtitle="Top 5 and Bottom 5",
       x="Company",
       y="Average Gross Margin") +
  theme(
    legend.position= "none"
  )
```



```r
rm(gross_margin,top_gross_margin,bottom_gross_margin,top_bottom_gross_margrin)
```

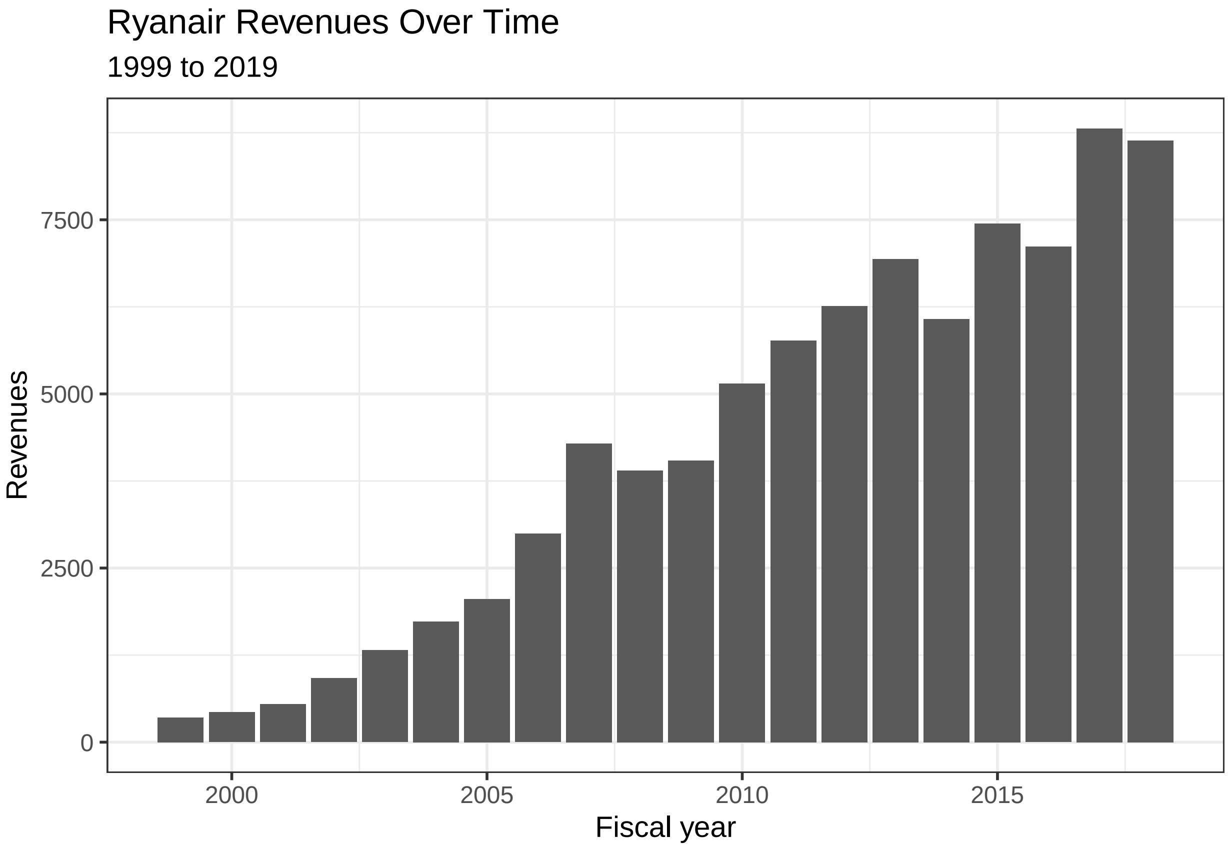
```r
# Industry revenues by year
airline_data %>%
  drop_na(sale) %>%
  group_by(fyear) %>%
  summarise(yearly_revenue = mean(sale)) %>%

ggplot(aes(x=fyear, y=yearly_revenue))+
  geom_col()+
  theme_bw()+
  labs(title="Airline Industry Revenues Over Time",
       subtitle = "1999 to 2019",
       x="Fiscal year",
       y="Revenues")+
  theme(
    legend.position= "none"
  )
```



Question 2

Select a Firm

```
#Chose Ryanair as Target Firm

#Ryanair Revenues
ryanair <- airline_data %>%
  drop_na(sale) %>%
  group_by(tic,conm, fyear) %>%
  filter(tic == "RYAAY")


ggplot(ryanair, aes(x=fyear, y=sale))+
  geom_col()+
  theme_bw()+
  labs(title="Ryanair Revenues Over Time",
       subtitle = "1999 to 2019",
       x="Fiscal year",
      y="Revenues")+
  theme(
    legend.position= "none"
  )
```



Question 3

Operating performance

```
# EBITDA margin = EBITDA/sales
# oibdp = Operating Income Before Depreciation (EBITDA)
airline_data <- airline_data %>%
  mutate(EBITDA_margin = oibdp/sale)
```

Credit / Liquidity

```
# quick ratio = (Current assets - inventories - prepaid assets)/current liabilities

airline_data <- airline_data %>%
  mutate(quick_ratio = (act- invt - xpp)/lct)
```

EV_to_EBITDA

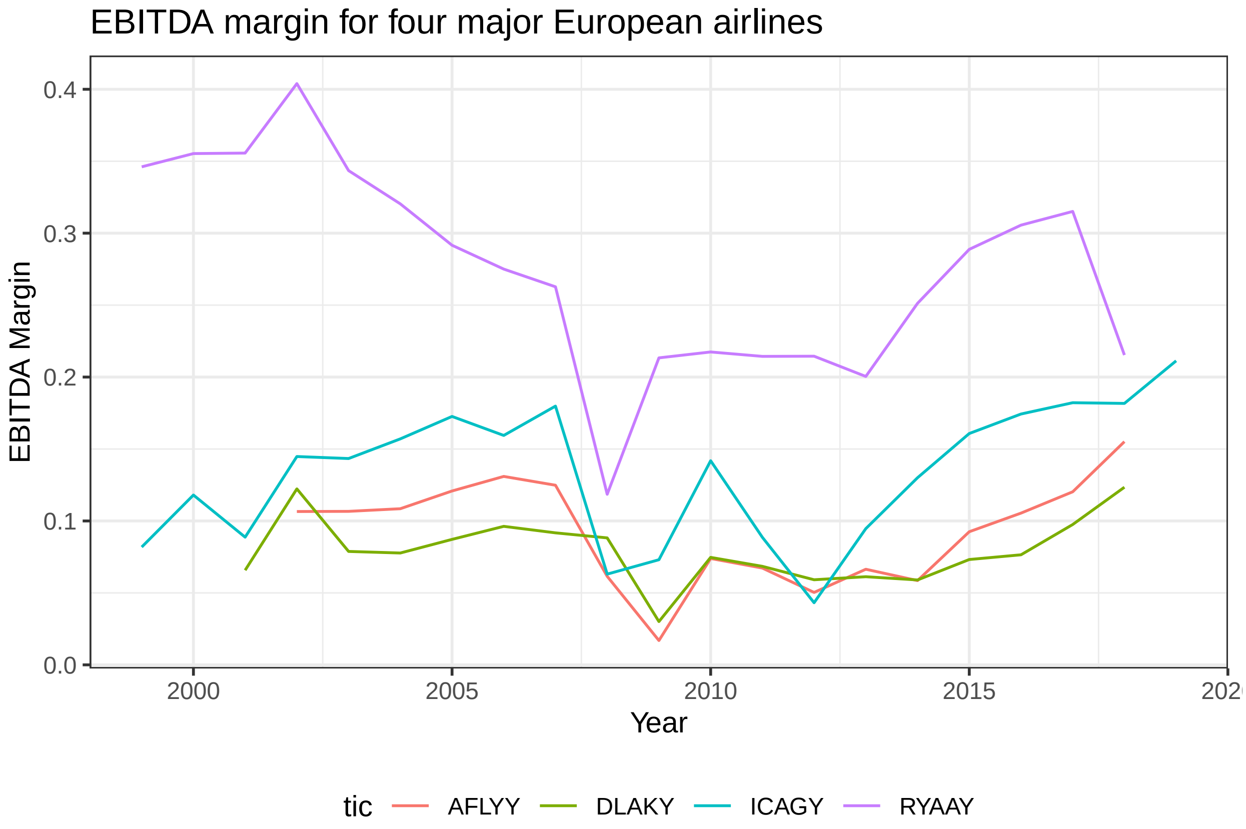
```
# Enterprise value = equity value + debt + preferred stock + noncontrolling interest - cash
# equity value = share price × fully diluted shares outstanding
# cash not available in the data set - we used che instead, because it represents cash + short-term inv

airline_data <- airline_data %>%
  mutate(EV_to_EBITDA = ((prcc_f*csho) + (dlc + dltt) + mib + pstk - che)/oibdp)
```

Question 4

```
# Based on the sector, size, and air routes, we decided to target the main airline players in Europe th

# DLAKY - Deutsche Lufthansa AG
# AFLYY - AIR FRANCE
# RYAAY - RYANAIR HOLDINGS PLC
# ICAGY - INTL CONSOL AIRLINES GROUP

# we need to run it again since the airline_data has three new variables

comparables_data <- airline_data %>%
  filter(tic %in% c("DLAKY", # 2001-2018
                 "AFLYY", # 2002-2018
                 "RYAAY", # 1999-2019
                 "ICAGY")) # 1999-2019
```

EBITDA margin Comparison

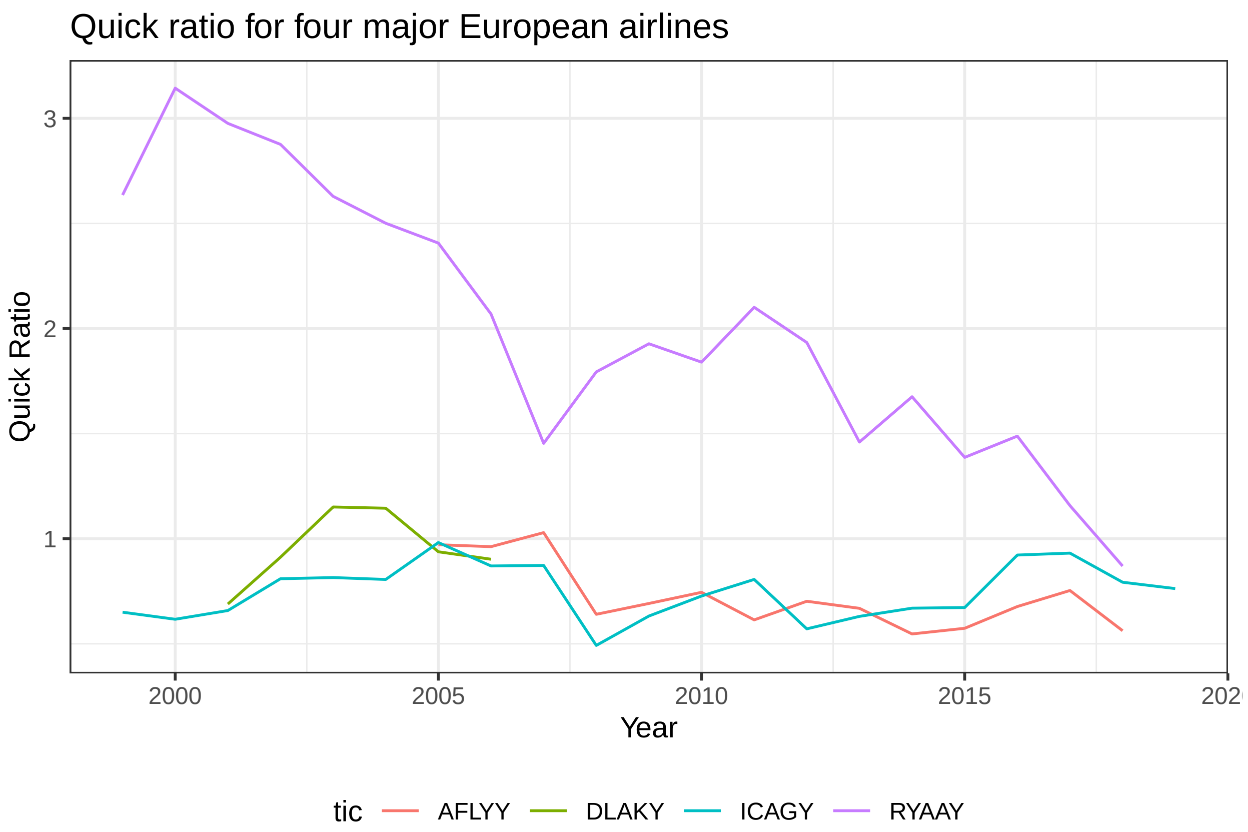
```
ggplot(comparables_data,aes(x=fyear,y=EBITDA_margin,colour=tic)) +
  geom_line() +
  theme_bw()+
  ggtitle("EBITDA margin for four major European airlines") +
  xlab("Year") + ylab("EBITDA Margin")+
  theme(legend.position="bottom")
```



Quick ratio Comparison

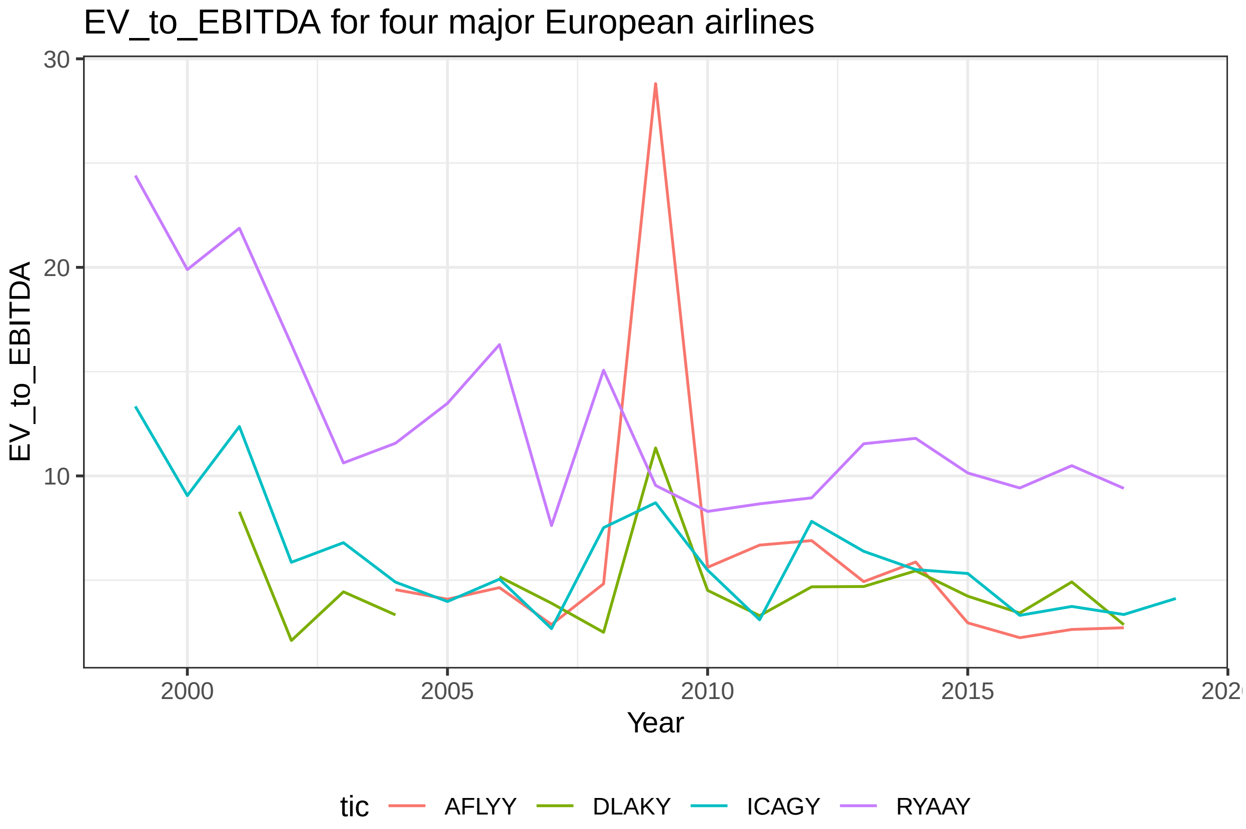
```
ggplot(comparables_data,aes(x=fyear,y=quick_ratio,colour=tic)) +
  geom_line() +
  theme_bw()+
  ggtitle("Quick ratio for four major European airlines") +
  xlab("Year") + ylab("Quick Ratio")+
  theme(legend.position="bottom")
```

```
## Warning: Removed 15 rows containing missing values (geom_path).
```



EV_to_EBITDA Comparison

```
ggplot(comparables_data,aes(x=fyear,y=EV_to_EBITDA,colour=tic)) +
  geom_line() +
  theme_bw()+
  ggtitle("EV_to_EBITDA for four major European airlines") +
  xlab("Year") + ylab("EV_to_EBITDA")+
  theme(legend.position="bottom")
```



Question 5

See Report

Part B

```
# import libraries
import os
import re
import numpy as np
import pandas as pd
import nltk
from nltk.corpus import stopwords
import matplotlib.pyplot as plt
from pandas.plotting import register_matplotlib_converters
import datetime
from datetime import date
from wordcloud import WordCloud, STOPWORDS
from collections import Counter
# Define exogenous variables
input_folder_path = 'assignment_2_files/letters/'
add_input_path = 'additional_input/cik_ticker.csv'
output_folder_path = 'results/'
# Configure the programme
if not os.path.exists(output_folder_path):
    os.makedirs(output_folder_path)
register_matplotlib_converters()
```

```
class Solve:
    '''Provides answers for Part B of Assignment 2'''
    def __init__(self, input_folder_path, add_input_path, output_folder_path):
        self.input_folder_path = input_folder_path
        self.df_ticker = pd.read_csv(add_input_path, sep = '|') # Load ticker and cik match info
        self.output_folder_path = output_folder_path
        self.letter_name_list = os.listdir(self.input_folder_path) # List all file names in a folder
        self.letter_name_list.remove('.DS_Store') # Remove the system file
        self.type_list = []
        self.date_trans_list = []
        self.cik_list = []
        self.df = pd.DataFrame()
        self.df_consolidated = pd.DataFrame()
        self.df_freq_firm_raw = pd.DataFrame()
        self.df_freq_firm = pd.DataFrame()
        self.files_dict = {}
        self.files_concat = ''
        self.wordcloud_input = None
        self.wc = None
        self.filtered_text_pre = None
        self.filtered_text = None
    def read_files(self):
        for filename in self.letter_name_list:
            with open(self.input_folder_path + filename, 'r') as file:
                self.files_dict[filename] = file.read()
                self.files_concat = self.files_concat + ' ' + self.files_dict[filename]
    def text_analytics(self):
        exclusion = set(stopwords.words('english'))
        mannual_exclusion = set(stopwords.words('english') + ['31', 'comments', 'company', 'please', 'fo
        self.filtered_text_pre = [i for i in self.files_concat.split() if i.lower() not in exclusion]
        self.filtered_text = [i for i in self.files_concat.split() if i.lower() not in mannual_exclusio
        self.wordcloud_input = dict(Counter(self.filtered_text).most_common(100))
        self.wc = WordCloud(background_color="white",width=2000,height=1500, max_words=50,relative_scal
        plt.figure(figsize = (16, 9))
        plt.imsave(fname = self.output_folder_path + 'Word_Cloud.jpeg', arr = self.wc, dpi = 300)
        plt.close()
        D = dict(Counter(self.filtered_text_pre).most_common(50))
        plt.figure(figsize=(25,10))
        plt.bar(range(len(D)), list(D.values()), align='center')
        plt.xticks(range(len(D)), list(D.keys()))
        plt.xticks(rotation=45)
        plt.xlabel('Words')
        plt.title('Words frequency')
        plt.ylabel('frequency')
        plt.savefig(self.output_folder_path + 'Word_Frequency.png', dpi = 300)
        plt.close()
    def overview(self):
        '''Obtains a list of letter types and transmission dates and stores in a dataframe'''
        # Extract letter types and transmission dates out of file names
        for letter_name in self.letter_name_list:
            self.type_list.append(
                ''.join(re.findall('[a-zA-Z]+', letter_name)).rstrip('txt')
                )
            match = re.search(r'\d{4}-\d{2}-\d{2}', letter_name)
            self.date_trans_list.append(
                datetime.datetime.strptime(match.group(), '%Y-%m-%d').date().strftime('%Y-%m-%d')
                )
        print('First three letter types:\n' + str(self.type_list[0:3]))
        print('First three transmission dates:\n' + str(self.date_trans_list[0:3]))
        # Store the extracted info in a dataframe
        self.df['letter_type'], self.df['date_trans'] = self.type_list, pd.to_datetime(self.date_trans_l
    def ticker_add(self):
        '''Extracts CIK numbers from file names and joins them with ticker information'''
        # Obtain a list of cik numbers for each letter
        [self.cik_list.append(letter_name[0:10].lstrip('0000')) for letter_name in self.letter_name_lis
        # Add cik number as a column to the consolidated dataframe
        self.df['cik'] = pd.Series(self.cik_list)
        # Join the ticker column with the consolidated dataframe after unifying column names
        self.df_ticker.rename(columns = {'CIK': 'cik'}, inplace = True)
        self.df_ticker = self.df_ticker[['cik', 'Name', 'Business']]
        self.df['cik'] = self.df['cik'].astype('int64')
        self.df_consolidated = self.df.join(self.df_ticker.set_index('cik'), on = 'cik')
        self.df_consolidated.reset_index(inplace = True)
        self.df_consolidated.drop(['index'], inplace = True, axis = 1)
        # Sort dataframes based on time of letters communicated
        self.df.sort_values(by = ['date_trans'], inplace = True)
    def analysis(self):
        self.df_freq_firm_raw = pd.DataFrame(self.df_consolidated['Name'].value_counts())
        self.df_freq_firm = self.df_freq_firm_raw.reset_index().rename(columns = {'index': 'Name', 'Name
    def data_export(self):
        # Produce a histogram of communication date frequency
        plt.figure(figsize = (16, 9))
        self.df_consolidated['date_trans'].hist(bins = 30)
        plt.title('Total Number of Letters Exchange over Time')
        plt.ylabel('Total Number of Letters)')
        plt.xlabel('First Transmission Date')
        plt.savefig(self.output_folder_path + 'Number_of_Letters_over_Time.png', dpi = 300)
        plt.close()
        # Save the extracted structured dataframe to disc
        self.df_consolidated.to_csv(output_folder_path + 'structured_data.csv', index = False)
    def data_export_2(self):
        # Produce a frequency chart for the number of letters each firm receives
        plt.figure(figsize = (16, 9))
        self.df_freq_firm_raw.plot(kind = 'barh', figsize = (20,9))
        plt.savefig(self.output_folder_path + 'Frequency_Chart_Firm.jpeg', dpi = 300)
        plt.close()
    def data_export_3(self):
        plt.figure(figsize = (16, 9))
        self.df_consolidated.loc[self.df_consolidated['cik'] == 27904, 'date_trans'].hist(bins = 30)
        plt.savefig(self.output_folder_path + 'Letter_Number_Delta.jpeg', dpi = 300)
        plt.close()
    def exec(self):
        self.read_files()
        self.text_analytics()
        self.overview()
        self.ticker_add()
        self.ticker_add()
        self.analysis()
        self.data_export()
        self.data_export_2()
        self.data_export_3()
        print('First 10 rows of the consolidated dataframe:')
        print(self.df_consolidated)
        print('Frequency table of letters communicated based on firms:')
        print(self.df_freq_firm)
```

```
# Execution code
obj = Solve(input_folder_path, add_input_path, output_folder_path)
obj.exec()
```

```
## First three letter types:
## ['UPLOAD', 'UPLOAD', 'CORRESP']
## First three transmission dates:
## ['2012-05-11', '2014-05-06', '2006-03-31']
## First 10 rows of the consolidated dataframe:
##     letter_type date_trans      cik                              Name Business
## 0        UPLOAD 2005-11-02   894081  Air Transport Services Group Inc       OH
## 1        UPLOAD 2005-11-16   319687               United Airlines Inc       IL
## 2       CORRESP 2005-11-21   319687               United Airlines Inc       IL
## 3       CORRESP 2005-11-22   894081  Air Transport Services Group Inc       OH
## 4        UPLOAD 2005-11-22    27904               Delta Air Lines Inc       GA
## 5        UPLOAD 2005-12-01   319687               United Airlines Inc       IL
## 6        UPLOAD 2005-12-02   894081  Air Transport Services Group Inc       OH
## 7       CORRESP 2005-12-07    27904               Delta Air Lines Inc       GA
## 8        UPLOAD 2005-12-14    27904               Delta Air Lines Inc       GA
## 9       CORRESP 2005-12-20    27904               Delta Air Lines Inc       GA
## 10       UPLOAD 2005-12-22    27904               Delta Air Lines Inc       GA
## 11       UPLOAD 2006-01-24   810332                Mesa Air Group Inc       AZ
## 12      CORRESP 2006-01-27   810332                Mesa Air Group Inc       AZ
## 13       UPLOAD 2006-02-13   810332                Mesa Air Group Inc       AZ
## 14       UPLOAD 2006-03-10   948845             Midwest Air Group Inc       WI
## 15      CORRESP 2006-03-31   948845             Midwest Air Group Inc       WI
## 16       UPLOAD 2006-04-18   948845             Midwest Air Group Inc       WI
## 17      CORRESP 2006-05-05   948845             Midwest Air Group Inc       WI
## 18      CORRESP 2006-05-19  1158463              Jetblue Airways Corp       NY
## 19      CORRESP 2006-05-31     6201       American Airlines Group Inc       TX
## 20      CORRESP 2006-05-31  1159154     Republic Airways Holdings Inc       IN
## 21      CORRESP 2006-06-06  1159154     Republic Airways Holdings Inc       IN
## 22       UPLOAD 2006-06-07  1159154     Republic Airways Holdings Inc       IN
## 23       UPLOAD 2006-06-07   948846              Airtran Holdings Inc       FL
## 24       UPLOAD 2006-06-22  1172222             Hawaiian Holdings Inc       HI
## 25      CORRESP 2006-06-23   948846              Airtran Holdings Inc       FL
## 26       UPLOAD 2006-06-29   948846              Airtran Holdings Inc       FL
## 27      CORRESP 2006-07-05  1172222             Hawaiian Holdings Inc       HI
## 28      CORRESP 2006-07-17     6201       American Airlines Group Inc       TX
## 29       UPLOAD 2006-07-20  1166291            Pinnacle Airlines Corp       TN
## ..          ...        ...      ...                               ...      ...
## 438      UPLOAD 2018-07-12   319687               United Airlines Inc       IL
## 439      UPLOAD 2018-07-12   100517   United Continental Holdings Inc       IL
## 440      UPLOAD 2018-07-16    27904               Delta Air Lines Inc       GA
## 441     CORRESP 2018-07-23   319687               United Airlines Inc       IL
## 442     CORRESP 2018-07-23   100517   United Continental Holdings Inc       IL
## 443      UPLOAD 2018-07-24   766421              Alaska Air Group Inc       WA
## 444     CORRESP 2018-08-01   766421              Alaska Air Group Inc       WA
## 445     CORRESP 2018-08-02    27904               Delta Air Lines Inc       GA
## 446     CORRESP 2018-08-14    27904               Delta Air Lines Inc       GA
## 447      UPLOAD 2018-08-29   100517   United Continental Holdings Inc       IL
## 448      UPLOAD 2018-08-29    27904               Delta Air Lines Inc       GA
## 449      UPLOAD 2018-08-29   319687               United Airlines Inc       IL
## 450     CORRESP 2018-09-05   319687               United Airlines Inc       IL
## 451     CORRESP 2018-09-05   100517   United Continental Holdings Inc       IL
## 452      UPLOAD 2018-09-07   766421              Alaska Air Group Inc       WA
## 453      UPLOAD 2018-09-07   100517   United Continental Holdings Inc       IL
## 454      UPLOAD 2018-09-07   319687               United Airlines Inc       IL
## 455      UPLOAD 2018-12-04   793733                       Skywest Inc       UT
## 456     CORRESP 2018-12-17   793733                       Skywest Inc       UT
## 457      UPLOAD 2019-01-31   793733                       Skywest Inc       UT
## 458     CORRESP 2019-03-28   894081  Air Transport Services Group Inc       OH
## 459      UPLOAD 2019-04-01  1158463              Jetblue Airways Corp       NY
## 460     CORRESP 2019-04-12  1158463              Jetblue Airways Corp       NY
## 461      UPLOAD 2019-04-22  1158463              Jetblue Airways Corp       NY
## 462      UPLOAD 2019-05-02    27904               Delta Air Lines Inc       GA
## 463     CORRESP 2019-05-15    27904               Delta Air Lines Inc       GA
## 464      UPLOAD 2019-06-10    27904               Delta Air Lines Inc       GA
## 465      UPLOAD 2019-08-15   894081  Air Transport Services Group Inc       OH
## 466     CORRESP 2019-08-23   894081  Air Transport Services Group Inc       OH
## 467      UPLOAD 2019-08-29   894081  Air Transport Services Group Inc       OH
##
## [468 rows x 5 columns]
## Frequency table of letters communicated based on firms:
##                                    Name  Frequency
## 0                   Delta Air Lines Inc         49
## 1       United Continental Holdings Inc         38
## 2                 Hawaiian Holdings Inc         33
## 3                  Jetblue Airways Corp         33
## 4           American Airlines Group Inc         32
## 5                   United Airlines Inc         29
## 6                 Southwest Airlines Co         26
## 7                           Skywest Inc         22
## 8      Air Transport Services Group Inc         21
## 9                  Alaska Air Group Inc         19
## 10                  Allegiant Travel Co         17
## 11        Republic Airways Holdings Inc         17
## 12               Pinnacle Airlines Corp         16
## 13                 Baltia Air Lines Inc         16
## 14                 Airtran Holdings Inc         12
## 15             Great Lakes Aviation LTD         10
## 16                 US Airways Group Inc         10
## 17              Expressjet Holdings Inc          5
## 18                Midwest Air Group Inc          4
## 19  Gulfstream International Group Inc          3
## 20       Frontier Airlines Holdings Inc          3
## 21                  Alas Aviation Corp          3
## 22                  Mesa Air Group Inc          3
## 23                 Spirit Airlines Inc          3
## 24             Northwest Airlines Corp          2
```