

Flights 2010-2020 version 1.1

Yulia Tyukhova

January 17, 2023

Introduction

Below I explore my cleaned-up 2010-2020 flight data set that contains personal and business flights. The data set contains 5 variables - **Date**, **Departure**, **Destination**, **PriceDOLLARS**, and **Airline** - displayed as a heading with 329 observations. Code below is focused on basic visualization of different patterns such as, for example, the cheapest day to fly averaged across 10 years and other variables based on 1 person's data.

Data Preparation

Load all packages

```
## For making data manipulation easier
library(dplyr)
## For creating simple tables
library(knitr)
## For creating graphics
library(ggplot2)
## For helping build common complex tables and manipulate table styles
library(kableExtra)
## For arranging multiple grid-based plots on a page
library(gridExtra)
```

Load the data set

```
Sys.setlocale("LC_ALL", "C")
```

```
## [1] "C"
```

```
## Load data
df<-read.csv("FlightsData_2.csv",header=TRUE,as.is=TRUE,
             na.strings=",")
data<-df
str(data)
```

```
## 'data.frame':   329 obs. of  5 variables:
## $ Date          : chr  "8/4/2010" "8/7/2010" "8/7/2010" "10/7/2010" ...
## $ Departure     : chr  "Moscow, Russia (DME)" "Washington DC (IAD)" "Chicago, IL (ORD)" "Omaha, NE (O)" ...
## $ Destination  : chr  "Washington DC (IAD)" "Chicago, IL (ORD)" "Omaha, NE (OMA)" "Denver, CO (DEN)" ...
## $ PriceDOLLARS : num  NA NA NA 152 NA ...
## $ Airline       : chr  "UA" "UA" "UA" "UA" ...
```

```
## Read in date correctly
Date<-as.Date(df$Date,"%m/%d/%Y")
## Assigning new data column to the data instead of the old format date
data$Date<-Date
## Make sure I have 329 rows of data without NAs in Date
## sum(!is.na(data$Date))
## Add weekdays column from the Date
data<-mutate(data,weekday=weekdays(data$Date))
## head(data)
## Split Date into 3 columns - year, month, day
data2<-data.frame(date=data$Date,
                  year=as.numeric(format(data$Date,format="%Y")),
                  month=as.numeric(format(data$Date,format="%m")),
                  day=as.numeric(format(data$Date,format="%d")))
## Combine 2 data sets and get rid of the extra date column
data3<-cbind(data,data2)
data3<-mutate(data3,date=NULL)
tail(data3)
```

```
##           Date           Departure           Destination PriceDOLLARS Airline
## 324 2019-12-29   San Jose, CA (SJC)   Las Vegas, NV (LAS)           NA      SW
## 325 2020-01-03   Las Vegas, NV (LAS)   San Jose, CA (SJC)           5.60      SW
## 326 2020-02-20   San Jose, CA (SJC)   Burbank, CA (BUR)          367.46      SW
## 327 2020-02-20   Burbank, CA (BUR)    Denver, CO (DEN)           NA      SW
## 328 2020-02-22   Denver, CO (DEN)   Los Angeles, CA (LAX)        NA      SW
## 329 2020-02-22  Los Angeles, CA (LAX)   San Jose, CA (SJC)           NA      SW
##      weekday year month day
## 324   Sunday 2019   12   29
## 325   Friday 2020    1    3
## 326 Thursday 2020    2   20
## 327 Thursday 2020    2   20
## 328 Saturday 2020    2   22
## 329 Saturday 2020    2   22
```

```
# Sort the number of flights per weekday from the smallest to the largest
sort(table(data3$weekday))
```

```
##
##   Saturday   Monday Wednesday   Friday Thursday   Tuesday   Sunday
##          39         40         44         46         49         49         62
```

```
# Make sure I didn't lose data - 329 observations
sum(table(data3$weekday))
```

```
## [1] 329
```

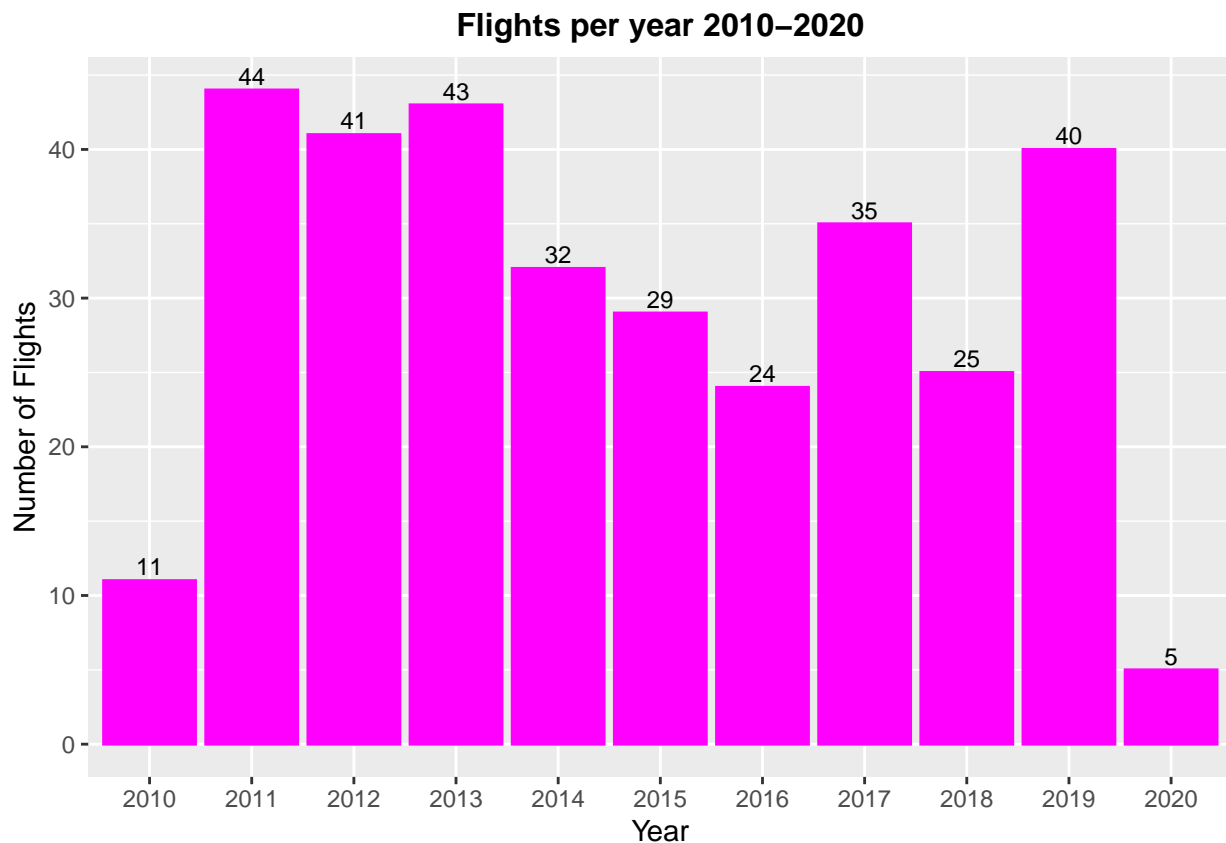
The number of flights per year

```
yeargr<-group_by(data3,year)
by_year<-summarize(yeargr,count=n())
```

```
dy<-as.data.frame(by_year)
names(dy)<-c("year","flights")
dy<-na.omit(dy)
dy
```

```
##   year flights
## 1  2010      11
## 2  2011      44
## 3  2012      41
## 4  2013      43
## 5  2014      32
## 6  2015      29
## 7  2016      24
## 8  2017      35
## 9  2018      25
## 10 2019      40
## 11 2020       5
```

```
g1<-ggplot(data=dy, aes(x=factor(year),y=flights))+
  geom_bar(stat="identity",color="magenta",fill="magenta")+
  ggtitle(label="Flights per year 2010-2020")+
  labs(x="Year",y="Number of Flights")+
  theme(plot.title = element_text(size = 12,hjust=0.5,face="bold"))+
  geom_text(aes(label=flights), vjust=-0.3, size=3)
g1
```



The number of flights exceeded or were equal to 40 in **2011, 2012, 2013 and 2019**.

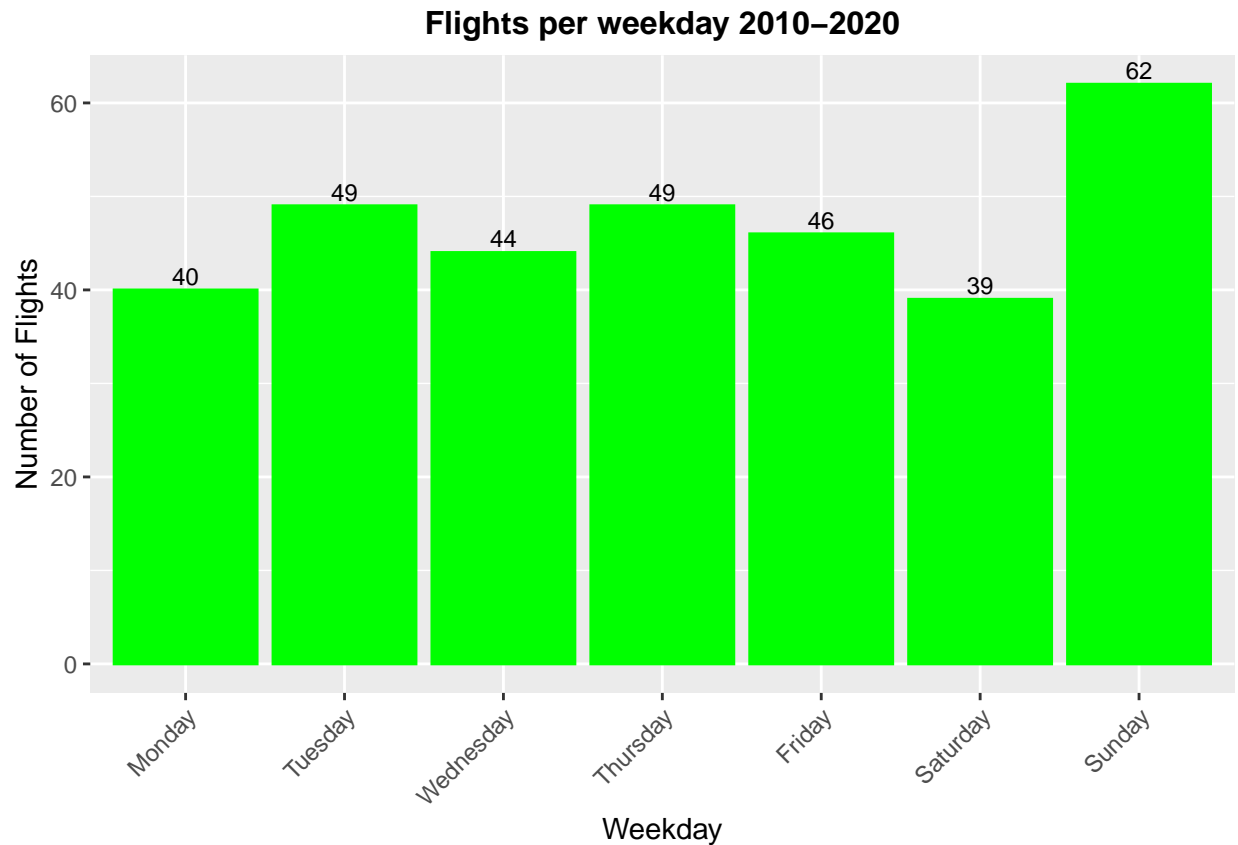
The number of flights per weekday

```
daygr<-group_by(data3,weekday)
by_day<-summarize(daygr,count=n())
dday<-as.data.frame(by_day)
names(dday)<-c("weekday","flights")
dday<-na.omit(dday)
dday
```

```
##      weekday flights
## 1      Friday      46
## 2      Monday      40
## 3     Saturday      39
## 4       Sunday      62
## 5    Thursday      49
## 6     Tuesday      49
## 7   Wednesday      44
```

```
dday$weekday<-factor(dday$weekday,levels=c("Monday","Tuesday", "Wednesday", "Thursday",
                                             "Friday", "Saturday", "Sunday"))
```

```
g2<-ggplot(data=dday, aes(x=weekday,y=flights))+
  geom_bar(stat="identity",color="green",fill="green")+
  ggtitle(label="Flights per weekday 2010-2020")+
  labs(x="Weekday",y="Number of Flights")+
  theme(axis.text.x = element_text(angle = 45, hjust = 1,vjust=1),
        plot.title = element_text(size = 12,hjust=0.5,face="bold"))+
  geom_text(aes(label=flights), vjust=-0.3, size=3)
g2
```



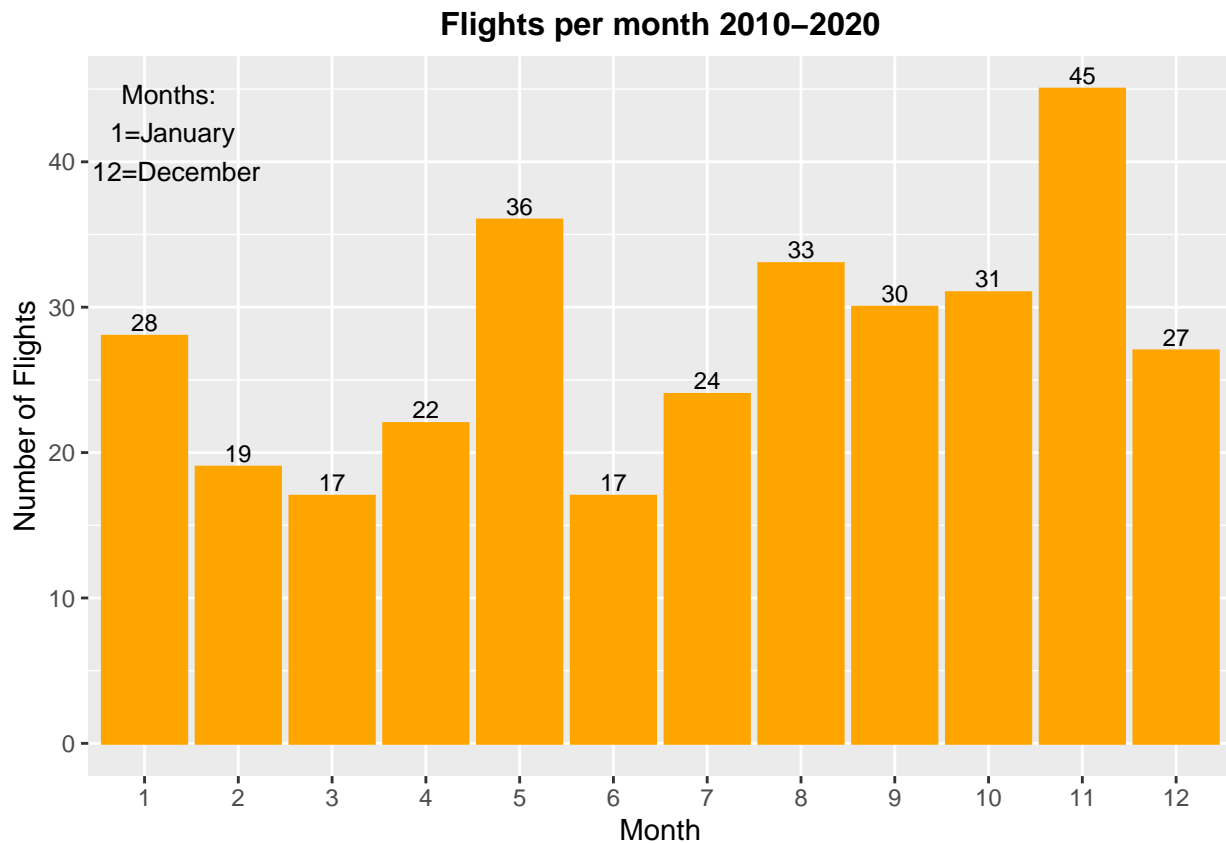
The majority of the flights were taken on **Sunday**. Otherwise, it is quite even across the entire week.

The number of flights per month

```
mgr<-group_by(data3,month)
by_month<-summarize(mgr,count=n())
dm<-as.data.frame(by_month)
names(dm)<-c("month","flights")
dm<-na.omit(dm)
dm
```

```
##   month flights
## 1     1      28
## 2     2      19
## 3     3      17
## 4     4      22
## 5     5      36
## 6     6      17
## 7     7      24
## 8     8      33
## 9     9      30
## 10    10      31
## 11    11      45
## 12    12      27
```

```
## Factor in x in aes allows to display x axis at discrete levels
g3<-ggplot(data=dm, aes(x=factor(month),y=flights))+
  geom_bar(stat="identity",color="orange",fill="orange")+
  ggtitle(label="Flights per month 2010-2020")+
  labs(x="Month",y="Number of Flights")+
  theme(plot.title = element_text(size = 12,hjust=0.5,face="bold"))+
  geom_text(aes(label=flights), vjust=-0.3, size=3)+
  annotate("text", x = 1.3, y = 42, label = "Months: \n 1=January \n 12=December",
          col = "black", size = 3.5)
g3
```



In 2010-2020 the most popular month to travel or the highest number of connections was in **November** with 45 flights, then **May** with 36, and **August** with 33.

Flights per airlines

First, I had to display all unique names for airlines I used in my data set. Any empty cells were replaced with NAs. Then I combined similar names - for example, SW and Southwest mean the same, therefore, they were combined.

```
unique(data3$Airline)
```

```
## [1] "UA" "DL"
## [3] "US Airways" "LUFTHANSA"
```

```
## [5] "Austrian"           "Air France"
## [7] "LOT Polish Airlines" "AirFrance"
## [9] "Southwest"          "AA"
## [11] "FinnAir"            "Tyrolean Airlines for Lufthansa"
## [13] "Swiss International Air Lines" "Air Berlin"
## [15] "AirTran"            "Lufthansa"
## [17] "SW"                 "Alaska"
## [19] "Jet Blue"           "Hawaiian"
## [21] "Aeroflot"           "Air Canada"
## [23] "Austrian Airlines"  "Vueling"
## [25] "Laudamotion"        "Swiss"
```

```
data3$Airline[data3$Airline==""]<-NA
data3$Airline[data3$Airline=="Southwest"]<-"SW"
data3$Airline[data3$Airline=="Swiss International Air Lines"]<-"Swiss"
data3$Airline[data3$Airline=="LUFTHANSA"]<-"Lufthansa"
data3$Airline[data3$Airline=="Austrian Airlines"]<-"Austrian"
data3$Airline[data3$Airline=="AirFrance"]<-"Air France"
data3$Airline[data3$Airline=="Tyrolean Airlines for Lufthansa"]<-"Tyrolean"
data3$Airline[data3$Airline=="LOT Polish Airlines"]<-"LOT Polish"
unique(data3$Airline)
```

```
## [1] "UA"           "DL"           "US Airways"   "Lufthansa"   "Austrian"
## [6] "Air France"   "LOT Polish"   "SW"           "AA"           "FinnAir"
## [11] "Tyrolean"     "Swiss"        "Air Berlin"   "AirTran"     "Alaska"
## [16] "Jet Blue"     "Hawaiian"     "Aeroflot"     "Air Canada"  "Vueling"
## [21] "Laudamotion"
```

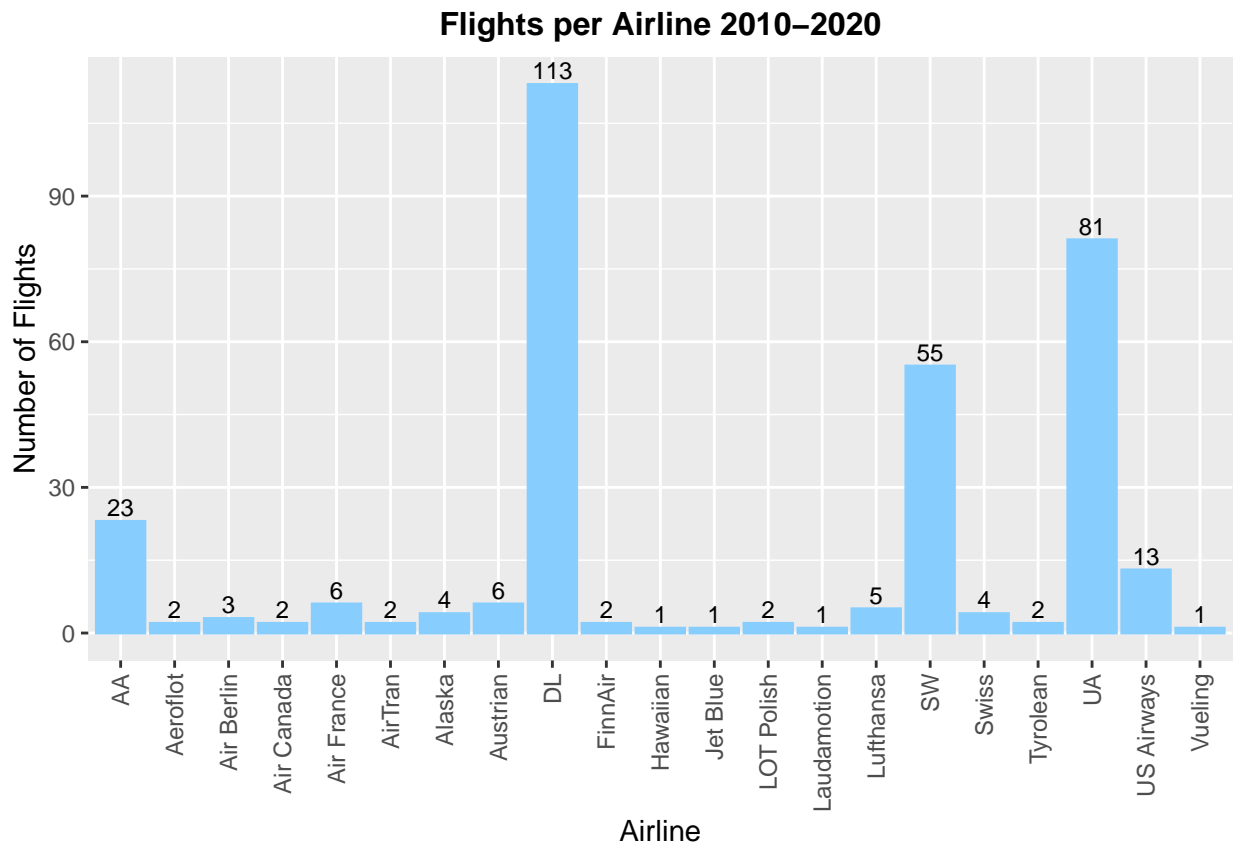
Now I have 21 unique airline names.

```
agr<-group_by(data3,Airline)
by_airline<-summarize(agr,count=n())
da<-as.data.frame(by_airline)
names(da)<-c("airline","flights")
da<-na.omit(da)
da
```

```
##      airline flights
## 1      AA         23
## 2  Aeroflot         2
## 3 Air Berlin         3
## 4 Air Canada         2
## 5 Air France         6
## 6   AirTran         2
## 7   Alaska          4
## 8   Austrian         6
## 9      DL        113
## 10  FinnAir         2
## 11 Hawaiian         1
## 12  Jet Blue         1
## 13 LOT Polish         2
## 14 Laudamotion        1
## 15 Lufthansa         5
```

```
## 16      SW      55
## 17    Swiss      4
## 18  Tyrolean      2
## 19      UA     81
## 20 US Airways    13
## 21    Vueling      1
```

```
g4<-ggplot(data=da, aes(x=airline,y=flights))+
  geom_bar(stat="identity",color="skyblue1",fill="skyblue1")+
  ggtitle(label="Flights per Airline 2010-2020")+
  labs(x="Airline",y="Number of Flights")+
  theme(axis.text.x = element_text(angle = 90, hjust = 1,vjust=0.5),
        plot.title = element_text(size = 12,hjust=0.5,face="bold"))+
  geom_text(aes(label=flights), vjust=-0.3, size=3)
g4
```



The overwhelming number of flights were taken on **Delta** 113, then **United** 81, **Southwest** 55, and **American** 23.

Most popular airline per year

```
aygr<-group_by(data3,year,Airline)
by_yearairline<-summarize(aygr,count=n())
```



```
## 'summarise()' has grouped output by 'year'. You can override using the
## '.groups' argument.
```

```
dya<-as.data.frame(by_yearairline)
names(dya)<-c("year","airline","flights")
dya<-na.omit(dya)
## dya
## The most popular airline by year
max_ay<-dya %>% group_by(year) %>% slice(which.max(flights))
maxay<-as.data.frame(max_ay)
maxay
```

```
##   year airline flights
## 1  2010      UA       11
## 2  2011      UA       28
## 3  2012      DL       19
## 4  2013      DL       14
## 5  2014      DL        7
## 6  2015      SW       10
## 7  2016      DL       14
## 8  2017      DL       18
## 9  2018      DL       16
## 10 2019      SW       13
## 11 2020      SW        5
```

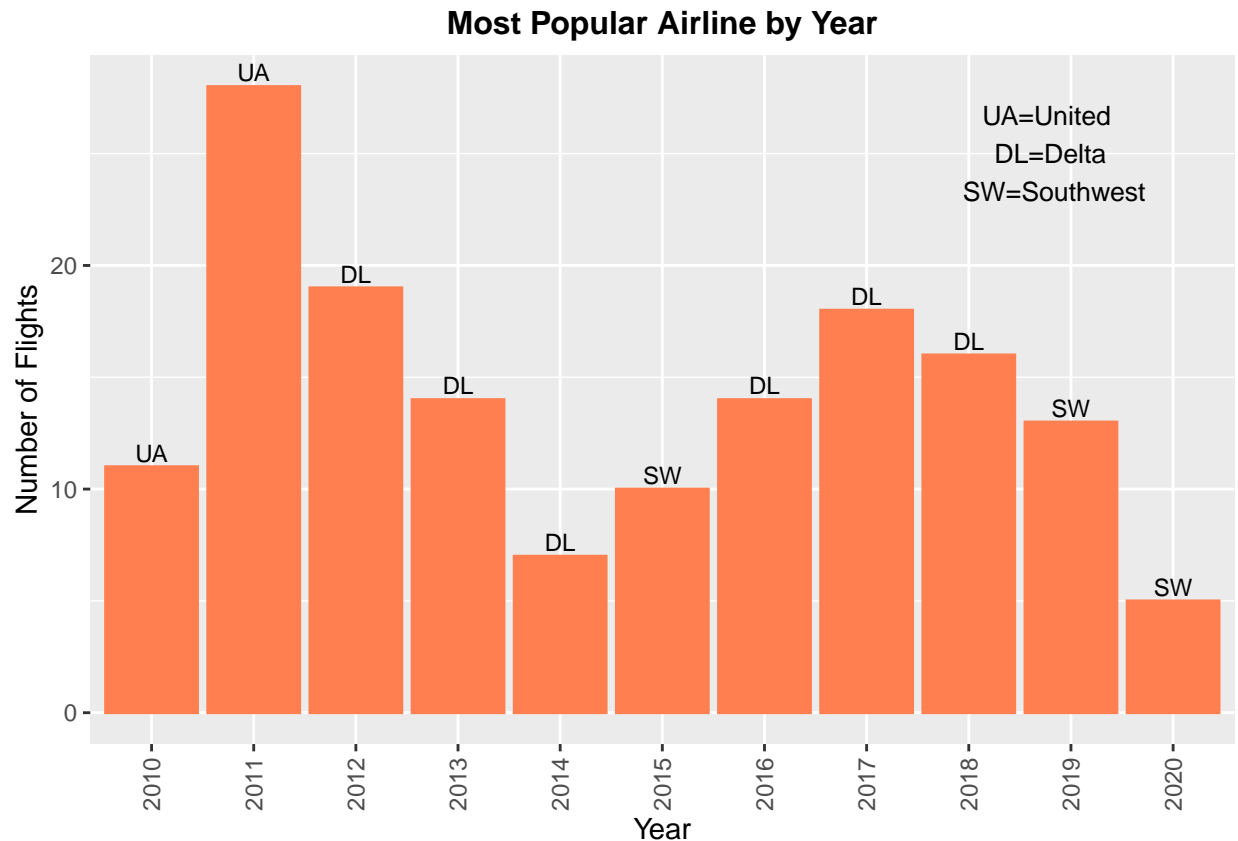
```
class(maxay$year)
```

```
## [1] "numeric"
```

```
## To display every year on x axis, year from numeric to factor
```

```
maxay$year <- as.factor(maxay$year)
```

```
g5<-ggplot(data=maxay, aes(x=year,y=flights))+
  geom_bar(stat="identity",color="coral",fill="coral")+
  ggtitle(label="Most Popular Airline by Year")+
  labs(x="Year",y="Number of Flights")+
  theme(axis.text.x = element_text(angle = 90, hjust = 1,vjust=0.5),
        plot.title = element_text(size = 12,hjust=0.5,face="bold"))+
  geom_text(aes(label=airline), vjust=-0.3, size=3)+
  annotate("text", x = 9.8, y = 25, label = "UA=United \n DL=Delta \n SW=Southwest",
         col = "black", size = 3.5)
g5
```



Delta was the most popular airline for 6 years, then **Southwest** for 3 years, and **United** for 2.

Most popular departure/layover and destination/layover airport

```
## The number of departure/layover locations in alphabetic order
dep<-sort(table(data3$Departure),decreasing=T)
sum(table(data3$Departure))
```

```
## [1] 329
```

```
class(dep)
```

```
## [1] "table"
```

```
dep<-as.data.frame(dep)
## Rename columns
names(dep)<-c("location","frequency")
## dep
dep1<-filter(dep,frequency>10)
dep1
```

```
##           location frequency
```

```
## 1      Atlanta, GA (ATL)      52
## 2      Omaha, NE (OMA)      52
## 3      Chicago, IL (ORD)     30
## 4      San Jose, CA (SJC)     30
## 5 San Francisco, CA (SFO)     13
## 6      Denver, CO (DEN)     12
```

```
## Most popular destination/layover airport
arr<-sort(table(data3$Destination),decreasing=T)
sum(table(data3$Destination))
```

```
## [1] 329
```

```
arr<-as.data.frame(arr)
## Rename columns
names(arr)<-c("location","frequency")
## arr
arr1<-filter(arr,frequency>10)
arr1
```

```
##           location frequency
## 1      Omaha, NE (OMA)      53
## 2      Atlanta, GA (ATL)     51
## 3      Chicago, IL (ORD)     30
## 4      San Jose, CA (SJC)     30
## 5 San Francisco, CA (SFO)     14
## 6      Denver, CO (DEN)     12
```

The most popular departure/layover was **Atlanta** and **Omaha** with 52 flights each followed by **Chicago** with 30. The most popular arrival/layover was **Omaha** with 53 flights and **Atlanta** with 51 followed by Chicago with 30.

Cost per weekday across 2010-2020

Cost analysis does not account for the number of flights/connections included for the given price. It is assumed that the price is valid for the total trip including the return ticket and disregarding the number of connections. Some personal flight prices might include tickets for two people.

```
## unique(data3$PriceDOLLARS)

## Info about the first leg of the trip that costs more than $1500
## filter(data3,PriceDOLLARS>1500)

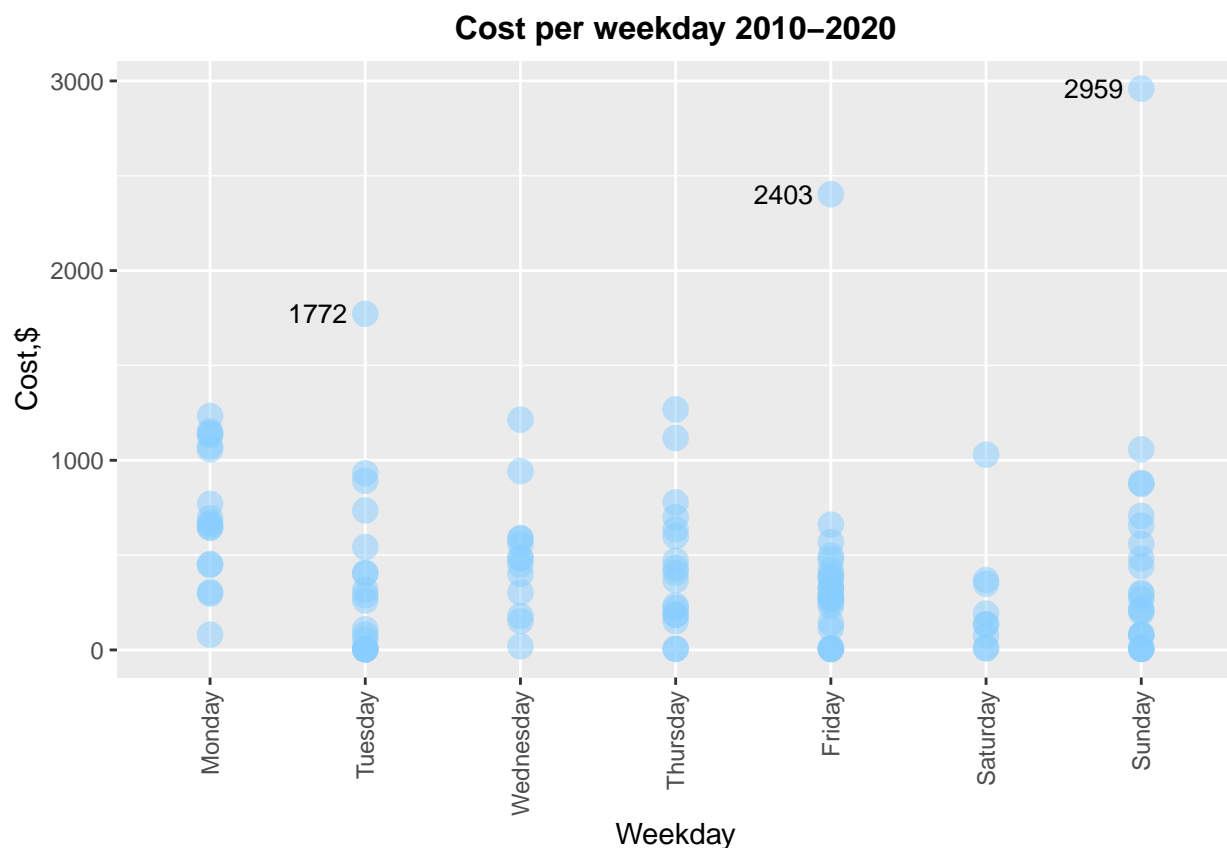
## To order the days of the week correctly
data3$weekday<-factor(data3$weekday,levels=c("Monday","Tuesday", "Wednesday",
                                              "Thursday","Friday", "Saturday",
                                              "Sunday"))

## Next two lines for labeling the subset of data in geom_text
top3<-subset(data3, PriceDOLLARS > 1500)
top3$PriceDOLLARS
```

```
## [1] 1772.34 2959.20 2402.64
```

```
g6<-ggplot(data=data3, aes(x=weekday,y=PriceDOLLARS))+
  geom_point(color="skyblue1",size=4, alpha=0.5)+
  ggtitle(label="Cost per weekday 2010-2020")+
  labs(x="Weekday",y="Cost,$")+
  theme(axis.text.x = element_text(angle = 90, hjust = 1,vjust=0.5),
        plot.title = element_text(size = 12,hjust=0.5,face="bold"))+
  geom_text(data=subset(data3, PriceDOLLARS > 1500),
            aes(x=weekday,y=PriceDOLLARS),label=round(top3$PriceDOLLARS), hjust=1.3, size=3.5)
g6
```

```
## Warning: Removed 208 rows containing missing values ('geom_point()').
```



```
## Prices over $1500 in the data set
##data3[which(data3$PriceDOLLARS>1500),]
## Row numbers
rownum<-as.numeric(rownames(data3[which(data3$PriceDOLLARS>1500),]))
## All data needed
trip1<-data[c(rownum[1],rownum[1]+1,rownum[1]+2,rownum[1]+3),]
trip2<-data[c(rownum[2],rownum[2]+1,rownum[2]+2),]
trip3<-data[c(rownum[3],rownum[3]+1,rownum[3]+2,rownum[3]+3,rownum[3]+4,rownum[3]+5),]
trip1
```

```
##      Date      Departure      Destination PriceDOLLARS Airline
```

```
## 33 2011-06-28      Omaha, NE (OMA)      Atlanta, GA (ATL)      1772.34      DL
## 34 2011-06-28      Atlanta, GA (ATL) Moscow, Russia (SVO)      NA      DL
## 35 2011-08-02 Moscow, Russia (SVO)      Atlanta, GA (ATL)      NA      DL
## 36 2011-08-02      Atlanta, GA (ATL)      Omaha, NE (OMA)      NA      DL
##      weekday
## 33 Tuesday
## 34 Tuesday
## 35 Tuesday
## 36 Tuesday
```

trip2

```
##      Date      Departure      Destination PriceDOLLARS Airline
## 67 2012-06-03      Omaha, NE (OMA) Minneapolis, MN (MSP)      2959.2      DL
## 68 2012-06-03 Minneapolis, MN (MSP)      New York, NY (JFK)      NA      DL
## 69 2012-06-03      New York, NY (JFK) Moscow, Russia (SVO)      NA      DL
##      weekday
## 67 Sunday
## 68 Sunday
## 69 Sunday
```

trip3

```
##      Date      Departure      Destination PriceDOLLARS
## 306 2019-05-24      San Francisco, CA (SFO)      Munich, Germany (MUC)      2402.64
## 307 2019-05-25      Munich, Germany (MUC)      Vienna, Austria (VIE)      NA
## 308 2019-05-27      Vienna, Austria (VIE)      Barcelona, Spain (BCN)      NA
## 309 2019-05-30      Barcelona, Spain (BCN)      Vienna, Austria (VIE)      NA
## 310 2019-06-09      Vienna, Austria (VIE) Zurich, Switzerland (ZRH)      NA
## 311 2019-06-09 Zurich, Switzerland (ZRH)      San Francisco, CA (SFO)      NA
##      Airline weekday
## 306      Lufthansa Friday
## 307 Austrian Airlines Saturday
## 308      Vueling Monday
## 309      Laudamotion Thursday
## 310      Swiss Sunday
## 311      Swiss Sunday
```

All three most expensive trips included travel abroad. The most expensive trip (\$2959) was a trip to another country with 2 connections and no return ticket. The second most expensive trip included travel abroad with 1 connection for 2 people (during this trip there was an additional trip from there). The third most expensive trip was a trip to another country with only 1 stop and the return ticket.

Cost by weekday averaged across 10 years

```
cost_by_weekday_gr<-group_by(data3,weekday,PriceDOLLARS)
by_weekdaycost<-summarize(cost_by_weekday_gr,count=n())
```

```
## 'summarise()' has grouped output by 'weekday'. You can override using the
## '.groups' argument.
```

```

weekdaycost<-as.data.frame(by_weekdaycost)
## The next line is very important for the final table not to have all values to be NAs
weekdaycost<-na.omit(weekdaycost)
average_weekdaycost<-weekdaycost%>%group_by(weekday)%>%summarize(round(mean(PriceDOLLARS),1))
average_weekdaycost<-as.data.frame(average_weekdaycost)
## Rename columns
names(average_weekdaycost)<-c("weekday","cost")
## Organize in the order of cost
average_weekdaycost<-arrange(average_weekdaycost,desc(cost))
##average_weekdaycost
kable(average_weekdaycost, caption="Cost of flight per weekday in the period of 2010-2020")%>%
  kable_styling(latex_options = "HOLD_position")

```

Table 1: Cost of flight per weekday in the period of 2010-2020

weekday	cost
Monday	732.9
Sunday	566.6
Wednesday	489.4
Thursday	485.0
Tuesday	425.5
Friday	388.2
Saturday	255.7

The most expensive day to fly in my data set was **Monday** with the price of **\$733** averaged across 10 years, airlines, and months.

Cost per year

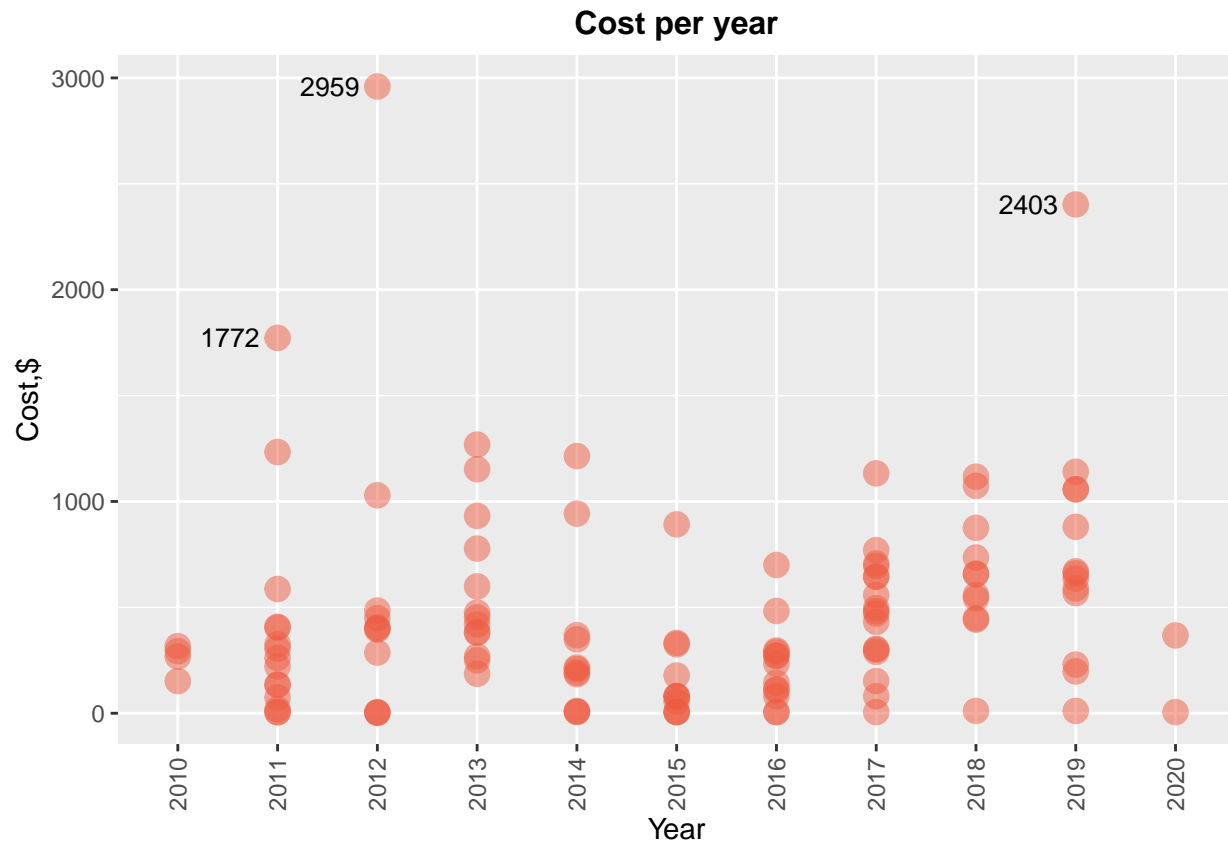
```

g7<-ggplot(data=data3, aes(x=factor(year),y=PriceDOLLARS))+
  geom_point(color="tomato2",size=4, alpha=0.5)+
  ggtitle(label="Cost per year")+
  labs(x="Year",y="Cost,$")+
  theme(axis.text.x = element_text(angle = 90, hjust = 1,vjust=0.5),
        plot.title = element_text(size = 12,hjust=0.5,face="bold"))+
  geom_text(data=subset(data3, PriceDOLLARS > 1500),
            aes(x=factor(year),y=PriceDOLLARS),label=round(top3$PriceDOLLARS),hjust=1.3, size=3.5)

g7

```

```
## Warning: Removed 208 rows containing missing values ('geom_point()').
```



```
## Cost by year averaged across weekdays
```

```
cost_by_year_gr<-group_by(data3,year,PriceDOLLARS)
by_yearcost<-summarize(cost_by_year_gr,count=n())
```

```
## 'summarise()' has grouped output by 'year'. You can override using the
## '.groups' argument.
```

```
yearcost<-as.data.frame(by_yearcost)
yearcost<-na.omit(yearcost)
average_yearcost<-yearcost%>%group_by(year)%>%summarize(round(mean(PriceDOLLARS),1))
average_yearcost<-as.data.frame(average_yearcost)
## Rename columns
names(average_yearcost)<-c("year","cost")
## Organize in the order of cost
average_yearcost<-arrange(average_yearcost,desc(cost))

kable(average_yearcost, caption="Cost of flight per year in the period of 2010-2020")%>%
  kable_styling(latex_options = "HOLD_position")
```

Table 2: Cost of flight per year in the period of 2010-2020

year	cost
2019	753.6
2018	647.3
2012	641.8
2013	597.0
2017	480.3
2011	420.5
2014	336.3
2015	267.0
2010	257.3
2016	249.8
2020	186.5

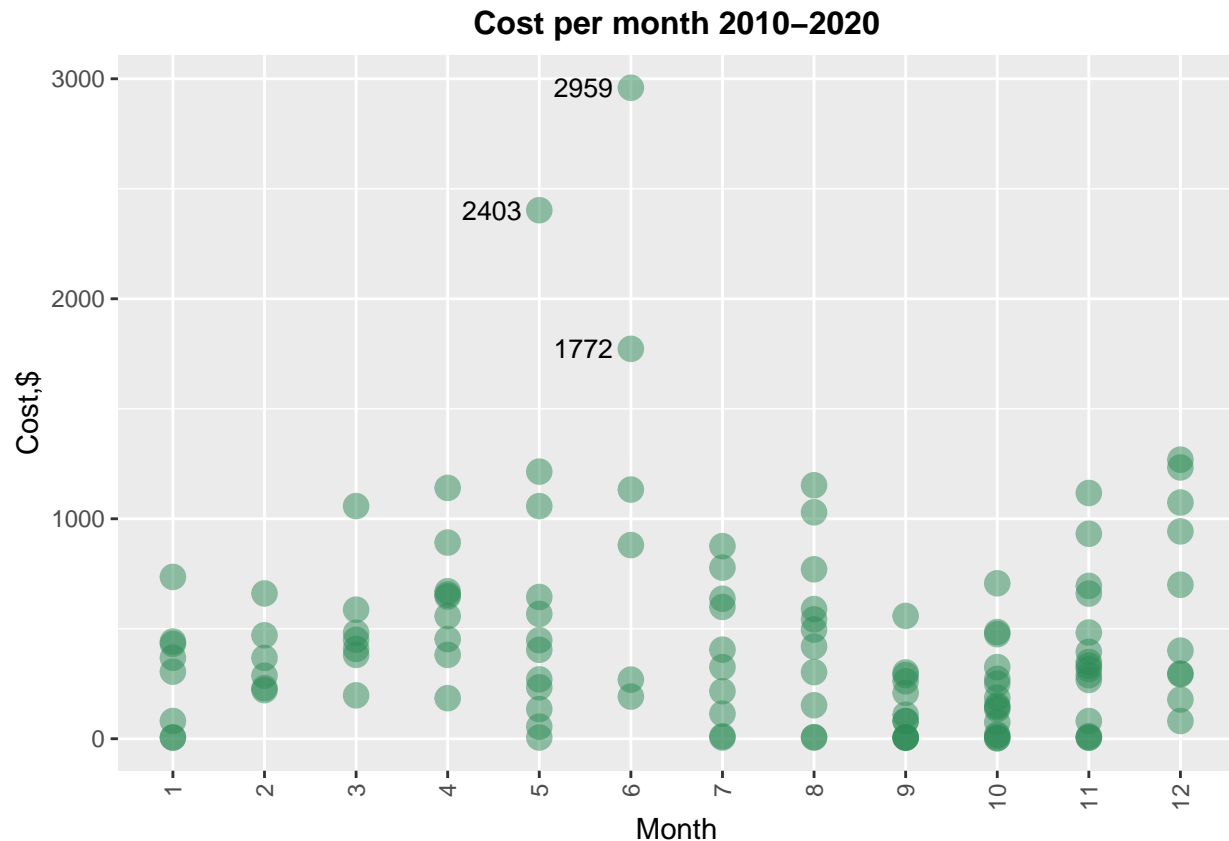
The most expensive year to fly in the data set was **2019** followed by 2018 and 2012. The cheapest year was **2020**. This is not surprising, since the number of flights I had taken before the pandemic started was only a handful.

Cost per month

```
g8<-ggplot(data=data3, aes(x=factor(month),y=PriceDOLLARS))+
  geom_point(color="seagreen4",size=4, alpha=0.5)+
  ggtitle(label="Cost per month 2010-2020")+
  labs(x="Month",y="Cost,$")+
  theme(axis.text.x = element_text(angle = 90, hjust = 1,vjust=0.5),
        plot.title = element_text(size = 12,hjust=0.5,face="bold"))+
  geom_text(data=subset(data3, PriceDOLLARS > 1500),
            aes(x=factor(month),y=PriceDOLLARS),label=round(top3$PriceDOLLARS),
              hjust=1.3, size=3.5)
```

g8

```
## Warning: Removed 208 rows containing missing values ('geom_point()').
```

```
## Cost by month averaged across 10 years
```

```
cost_by_month_gr<-group_by(data3,month,PriceDOLLARS)
by_monthcost<-summarize(cost_by_month_gr,count=n())
```

```
## 'summarise()' has grouped output by 'month'. You can override using the
## '.groups' argument.
```

```
monthcost<-as.data.frame(by_monthcost)
monthcost<-na.omit(monthcost)
average_monthcost<-monthcost%>%group_by(month)%>%summarize(round(mean(PriceDOLLARS),1))
average_monthcost<-as.data.frame(average_monthcost)
```

```
## Rename columns
```

```
names(average_monthcost)<-c("month","cost")
```

```
## Organize in the order of cost
```

```
average_monthcost<-arrange(average_monthcost,desc(cost))
```

```
kable(average_monthcost, caption="Cost of trip per month in the period of 2010-2020")%>%
  kable_styling(latex_options = "HOLD_position")
```

Table 3: Cost of trip per month in the period of 2010-2020

month	cost
6	1200.8
12	646.8
4	620.1
5	619.8
3	509.0
8	497.1
11	423.8
7	396.6
2	372.7
1	338.3
10	231.4
9	202.2

The most expensive month to fly in the period of 2010-2020 was **June**, and it was almost twice as expensive as the next one on the list - **December**. Perhaps, it was driven by two outliers - two out of three most expensive trips during the period of 2010-2020 took place in June. The cheapest month to fly was **September**.

Summary

- The number of flights exceeded or were equal to 40 in **2011, 2012, 2013 and 2019**.
- The majority of the flights were taken on **Sunday**. Otherwise, it is quite even across the entire week.
- In 2010-2020 the most popular month to travel or the highest number of connections was in **November** with 45 flights, then **May** with 36, and **August** with 33.
- The overwhelming number of flights were taken on **Delta** 113, then **United** 81, **Southwest** 55, and **American** 23.
- **Delta** was the most popular airline for 6 years, then **Southwest** for 3 years, and **United** for 2.
- The most popular departure/layover was **Atlanta** and **Omaha** with 52 flights each followed by **Chicago** with 30. The most popular arrival/layover was **Omaha** with 53 flights and **Atlanta** with 51 followed by Chicago with 30.
- The most expensive day to fly in my data set was **Monday** with the price of **\$733** averaged across 10 years, all airlines, and months.
- The most expensive year to fly in my data set was **2019** followed by 2018 and 2012. The cheapest year was **2020** with only handful number of flights because of the pandemic.
- The most expensive month to fly in the period 2010-2020 was **June**, and it was almost twice as expensive as the next one on the list - **December**. The cheapest month to fly was **September**.