**Assignment 2**

install.packages("hflights")

hf <- hflights::hflights

1. How Many Unique destinations from the origin.

a. Printing Origin and count of different destination

> hf %>% distinct( Origin, Dest) %>% group\_by(Origin) %>% summarise(Destination = n())

# A tibble: 2 x 2

Origin Destination

*<chr>* *<int>*

1 HOU 38

2 IAH 111

b. Printing Origin and different destination

> hf %>% distinct( Origin, Dest) %>% top\_n(10)

Selecting by Dest

Origin Dest

1 IAH TPA

2 IAH TUL

3 IAH TUS

4 HOU STL

5 HOU TPA

6 HOU TUL

7 IAH STL

8 IAH XNA

9 IAH TYS

1. IAH VPS
2. Which route has highest delay

> hf %>% group\_by(Origin, Dest) %>% summarise(MaxDepDelay = max(DepDelay, na.rm = T)) %>% arrange(desc(MaxDepDelay)) %>% head()

# A tibble: 6 x 3

# Groups: Origin [2]

Origin Dest MaxDepDelay

*<chr>* *<chr>* *<int>*

1 IAH HNL 981

2 IAH DFW 970

3 IAH ORD 931

4 IAH SFO 869

5 HOU DFW 803

6 IAH DEN 780

1. Monthly average delay per carrier

> hf %>% group\_by(Year,Month,UniqueCarrier) %>% summarise(DeptDelayTime = mean(DepDelay,na.rm=T))

# A tibble: 175 x 4

# Groups: Year, Month [?]

Year Month UniqueCarrier DeptDelayTime

*<int>* *<int>* *<chr>* *<dbl>*

1 2011 1 AA 5.03

2 2011 1 AS 1.45

3 2011 1 B6 13.5

4 2011 1 CO 8.98

5 2011 1 DL 8.03

6 2011 1 EV 21.7

7 2011 1 F9 6.31

8 2011 1 FL 3.94

9 2011 1 MQ 9.90

10 2011 1 OO 6.69

1. Unique carrier diverted count

> hf %>% filter(Diverted == 1) %>% group\_by(UniqueCarrier) %>% summarise(TotalDiversion = n())

# A tibble: 13 x 2

UniqueCarrier TotalDiversion

*<chr>* *<int>*

1 AA 6

2 AS 1

3 B6 4

4 CO 184

5 DL 8

6 EV 7

7 FL 7

8 MQ 9

9 OO 56

10 UA 5

11 US 6

12 WN 104

13 XE 252

1. Number of Cancellation (in each Carrier)

> hf %>% filter(Cancelled == 1) %>% group\_by(UniqueCarrier) %>% summarise(TotalCancellation = n())

# A tibble: 14 x 2

UniqueCarrier TotalCancellation

*<chr>* *<int>*

1 AA 60

2 B6 18

3 CO 475

4 DL 42

5 EV 76

6 F9 6

7 FL 21

8 MQ 135

9 OO 224

10 UA 34

11 US 46

12 WN 703

13 XE 1132

14 YV 1

1. Frequency of Cancellation ( Number of Cancellation on every month)

> hf %>% filter(Cancelled == 1 ) %>% group\_by(Year,Month) %>% summarise(TotalCancellation = sum(Cancelled,na.rm=T))

# A tibble: 12 x 3

# Groups: Year [?]

Year Month TotalCancellation

*<int>* *<int>* *<int>*

1 2011 1 209

2 2011 2 1108

3 2011 3 159

4 2011 4 251

5 2011 5 279

6 2011 6 169

7 2011 7 162

8 2011 8 182

9 2011 9 113

10 2011 10 112

11 2011 11 56

12 2011 12 173

1. Max and Min Hour of Travel(Minimum and Maximum Hour of Journey for each route)

> hf %>% group\_by(Origin,Dest) %>% summarise(MinumJourneyHr = min(ActualElapsedTime,na.rm=T), MaximumJourneyHr = max(ActualElapsedTime, na.rm = T))

# A tibble: 149 x 4

# Groups: Origin [?]

Origin Dest MinumJourneyHr MaximumJourneyHr

*<chr>* *<chr>* *<int>* *<int>*

1 HOU ABQ 103 212

2 HOU ATL 93 217

3 HOU AUS 34 95

4 HOU BHM 75 134

5 HOU BKG 74 98

6 HOU BNA 87 195

7 HOU BWI 137 247

8 HOU CHS 112 204

9 HOU CRP 37 89

10 HOU DAL 43 181

1. Number of flights based on hour

> hf %>% group\_by(DepTime%/%100) %>% summarise(TotalFlight = n())

# A tibble: 25 x 2

`DepTime%/%100` TotalFlight

*<dbl>* *<int>*

1 0 113

2 1 24

3 2 1

4 4 1

5 5 2393

6 6 5679

7 7 16236

8 8 10773

9 9 14352

10 10 17350

# ... with 15 more rows

1. Origin with Cancellation

> hf %>% filter(Cancelled == 1) %>% group\_by(Origin) %>% summarise(TotalCancellation = n())

# A tibble: 2 x 2

Origin TotalCancellation

*<chr>* *<int>*

1 HOU 868

2 IAH 2105

1. Average time between Arrival and departure time (for each Airport)
   1. Average Arrival delay time of each airport

> hf %>% group\_by(Origin) %>% summarise(DeptDelay = mean(DepDelay,na.rm=T)) %>% arrange(DeptDelay)

# A tibble: 2 x 2

Origin DeptDelay

*<chr>* *<dbl>*

1 IAH 8.44

2 HOU 12.8

* 1. Average departure delay time of each airport

> hf %>% group\_by(Dest) %>% summarise(ArrivalDelay = mean(ArrDelay,na.rm=T)) %>% arrange(ArrivalDelay)

# A tibble: 116 x 2

Dest ArrivalDelay

*<chr>* *<dbl>*

1 BKG -16.2

2 BFL -13.2

3 GRK -8.3

4 MTJ -0.451

5 HDN 0.0865

6 SNA 0.358

7 MIA 0.649

8 SJC 0.737

9 CLE 1.04

10 SLC 1.10

# ... with 106 more rows

1. Airport (Origin) with highest Taxi in time, Airport ( Departure) with lowest Taxi out time
   1. Highest Taxi in

> hf %>% group\_by(Origin) %>% summarise(LongestTaxiIn = max(TaxiIn,na.rm=T)) %>% arrange(desc(LongestTaxiIn)) %>% head(1)

# A tibble: 1 x 2

Origin LongestTaxiIn

*<chr>* *<int>*

1 IAH 165

* 1. Lowest Taxi out

> hf %>% group\_by(Dest) %>% summarise(ShortestTaxiOut = min(TaxiOut,na.rm=T)) %>% arrange(ShortestTaxiOut) %>% head(1)

# A tibble: 1 x 2

Dest ShortestTaxiOut

*<chr>* *<int>*

1 BNA 1

1. Average time between arrival and departure (delay for each carrier)
   1. Average arrival time delay of each carrier

> hf %>% group\_by(UniqueCarrier) %>% summarise(ArrivalDelay = mean(ArrDelay,na.rm=T)) %>% arrange(ArrivalDelay)

# A tibble: 15 x 2

UniqueCarrier ArrivalDelay

*<chr>* *<dbl>*

1 US -0.631

2 AA 0.892

3 FL 1.85

4 AS 3.19

5 YV 4.01

6 DL 6.08

7 CO 6.10

8 MQ 7.15

9 EV 7.26

10 WN 7.59

11 F9 7.67

12 XE 8.19

13 OO 8.69

14 B6 9.86

15 UA 10.5

* 1. Average departure time delay of each carrier

> hf %>% group\_by(UniqueCarrier) %>% summarise(DeptDelay = mean(DepDelay,na.rm=T)) %>% arrange(DeptDelay)

# A tibble: 15 x 2

UniqueCarrier DeptDelay

*<chr>* *<dbl>*

1 YV 1.54

2 US 1.62

3 AS 3.71

4 FL 4.72

5 F9 5.09

6 AA 6.39

7 XE 7.71

8 OO 8.89

9 CO 9.26

10 DL 9.37

11 MQ 11.1

12 EV 12.5

13 UA 12.9

14 B6 13.3

15 WN 13.5

1. Busiest hour / Peak hour (Top 4 peak hour with count of departure flights)

> hf %>% group\_by(DepTime%/%100) %>% summarise(TotalFlights = n()) %>% arrange(desc(TotalFlights)) %>% top\_n(4)

Selecting by TotalFlights

# A tibble: 4 x 2

`DepTime%/%100` TotalFlights

*<dbl>* *<int>*

1 19 19373

2 10 17350

3 14 17283

4 7 16236

1. Total distance covered (Total distance covered by each Carrier)

> hf %>% filter(Cancelled == 0) %>% group\_by(UniqueCarrier) %>% summarise(TotalDistanceCovered = sum(Distance))

# A tibble: 15 x 2

UniqueCarrier TotalDistanceCovered

*<chr>* *<int>*

1 AA 1543496

2 AS 684010

3 B6 966756

4 CO 76366034

5 DL 1880120

6 EV 1653703

7 F9 734439

8 FL 1451880

9 MQ 2943511

10 OO 13008934

11 UA 2402737

12 US 3961466

13 WN 27239670

14 XE 42444147

15 YV 73242

1. For a route, which air carrier completes in Quicker-Time

hf %>% group\_by(UniqueCarrier, Origin, Dest) %>% summarise(MinumJourneyHr = min(ActualElapsedTime,na.rm=T))

# A tibble: 243 x 4

# Groups: UniqueCarrier, Origin [?]

UniqueCarrier Origin Dest MinumJourneyHr

*<chr>* *<chr>* *<chr>* *<int>*

1 AA IAH DFW 49

2 AA IAH MIA 118

3 AS IAH SEA 241

4 B6 HOU JFK 165

5 CO IAH ABQ 111

6 CO IAH AGS 124

7 CO IAH ANC 384

8 CO IAH ATL 102

9 CO IAH AUS 36

10 CO IAH BHM 98