Tian Qiu
00265-35063

1. 26 variables.

   Categorical or qualitative variables: Year, Month, DayOfMonth, DayOfWeek, DepTime, ArrTime, UniqueCarrier, FlightNum, Origin, Dest.

   Quantitative or numeric variables: rest of the variable.

2. What is the standard deviation of Distance in this dataset?

   What is the range of distance in this dataset?

3. airline2008NovS <-
   read.delim("~/Desktop/STAT350/STAT350/Labs/Lab1/airline2008NovS.txt")

4. There are 3 missing values (NA) in the dataset.

   (a)

   cleaned <-airline2008NovS[complete.cases(airline2008NovS),]

   dim(cleaned)

   View(cleaned)

   write.table(helicon_cleaned,file="airCleaned.txt",quote=F,

   row.names=F, sep="\t")

   (b) 9997 observations.

(c) ~/Desktop/STAT350/STAT350/Labs/Lab1

5. (a)

```
airline2008NovS_new <- airline2008NovS_cleaned

airline2008NovS_new$Dest <-
as.character(airline2008NovS_new$Dest)

airline2008NovS_new$Dest[airline2008NovS_new$Dest==
"ATL"]<-"Atlanta"

airline2008NovS_new$Dest[airline2008NovS_new$Dest==
"CHS"]<-"CharlestonAFB"

airline2008NovS_new$Dest[airline2008NovS_new$Dest==
"DFW"]<-"DallasFtWorth"

airline2008NovS_new$Dest[airline2008NovS_new$Dest==
"MSP"]<-"MinneapolisStPaul"

View(airline2008NovS_new)
```

(b)

```
> airline2008NovS_new <- airline2008NovS_cleaned
> airline2008NovS_new$Dest <- as.character(airline2008NovS_new$Dest)
> airline2008NovS_new$Dest[airline2008NovS_new$Dest=="ATL"]<-"Atlanta"
> airline2008NovS_new$Dest[airline2008NovS_new$Dest=="CHS"]<-"CharlestonAFB"
> airline2008NovS_new$Dest[airline2008NovS_new$Dest=="DFW"]<-"DallasFtWorth"
> airline2008NovS_new$Dest[airline2008NovS_new$Dest=="MSP"]<-"MinneapolisStPaul"
> View(airline2008NovS_new)
> head(airline2008NovS_new)
  Year Month DayofMonth DayOfWeek DepTime CRSDepTime ArrTime CRSArrTime UniqueCarrier FlightNum TailNum ActualElapsedTime
1 2008    11          1         6    1001        913    1153       1125            9E      3747  86319E                112
2 2008    11          1         6    2116       2120    2345       2259            9E      2824  85339E                209
3 2008    11          1         6    1330       1235    1456       1413            9E      2070  91709E                 86
4 2008    11          1         6    1435       1400    1539       1518            9E      2024  91629E                 64
5 2008    11          1         6    1225       1211    1400       1320            9E      2024  91629E                 95
6 2008    11          1         6    1950       1850      17       2320            AA       614  N5BYAA                207
  CRSElapsedTime AirTime ArrDelay DepDelay Origin          Dest Distance TaxiIn TaxiOut CarrierDelay WeatherDelay NASDelay
1            132      99       28       48    DTW           SAV      706      4       9           28            0        0
2            159     130       46       -4    MSP           IDA      939      5      74            0            0       46
3             98      70       43       55    CVG           PHL      507      4      12            0            0        0
4             78      42       21       35    CHS       Atlanta      259     13       9            0            0        0
5             69      47       40       14    ATL CharlestonAFB      259     20      28           14            0       26
6            210     191       57       60    DFW           BOS     1562      5      11           14            0        0
  SecurityDelay LateAircraftDelay
1             0                 0
2             0                 0
3             0                43
4             0                21
5             0                 0
6             0                43
```

6. (9 points) We are going to see if the variable "ActualElapsedTime" can be calculated from other variables in the data set.

a. (3 pts.) Write down a mathematical equation to calculate "ActualElapsedTime" from “AirTime”, “TaxiIn” and “Taxiout”.

ActualElapsedTime = AirTime + TaxiIn +Taxiout

b. (3 pts.)

airline2008NovS_new$ActualElapsedTime = airline2008NovS_new$AirTime + airline2008NovS_new$TaxiIn +airline2008NovS_new$TaxiOut

head(airline2008NovS_new)

c. (3 pts.)

```
> airline2008NovS_new$ActualElapsedTime = airline2008NovS_new$AirTime + airline2008NovS_new$TaxiIn +airline2008NovS_new$TaxiOut
> head(airline2008NovS_new)
  Year Month DayofMonth DayOfWeek DepTime CRSDepTime ArrTime CRSArrTime UniqueCarrier FlightNum TailNum ActualElapsedTime
1 2008    11          1         6    1001        913    1153       1125            9E      3747  86319E                112
2 2008    11          1         6    2116       2120    2345       2259            9E      2824  85339E                209
3 2008    11          1         6    1330       1235    1456       1413            9E      2070  91709E                 86
4 2008    11          1         6    1435       1400    1539       1518            9E      2024  91629E                 64
5 2008    11          1         6    1225       1211    1400       1320            9E      2024  91629E                 95
6 2008    11          1         6    1950       1850      17       2320            AA       614  N5BYAA                207
  CRSElapsedTime AirTime ArrDelay DepDelay Origin          Dest Distance TaxiIn TaxiOut CarrierDelay WeatherDelay NASDelay
1            132      99       28       48    DTW           SAV      706      4       9           28            0        0
2            159     130       46       -4    MSP           IDA      939      5      74            0            0       46
3             98      70       43       55    CVG           PHL      507      4      12            0            0        0
4             78      42       21       35    CHS       Atlanta      259     13       9            0            0        0
5             69      47       40       14    ATL CharlestonAFB      259     20      28           14            0       26
6            210     191       57       60    DFW           BOS     1562      5      11           14            0        0
  SecurityDelay LateAircraftDelay
1             0                 0
2             0                 0
3             0                43
4             0                21
5             0                 0
6             0                43
> head(airline2008NovS_cleaned)
  Year Month DayofMonth DayOfWeek DepTime CRSDepTime ArrTime CRSArrTime UniqueCarrier FlightNum TailNum ActualElapsedTime
1 2008    11          1         6    1001        913    1153       1125            9E      3747  86319E                112
2 2008    11          1         6    2116       2120    2345       2259            9E      2824  85339E                209
3 2008    11          1         6    1330       1235    1456       1413            9E      2070  91709E                 86
4 2008    11          1         6    1435       1400    1539       1518            9E      2024  91629E                 64
5 2008    11          1         6    1225       1211    1400       1320            9E      2024  91629E                 95
6 2008    11          1         6    1950       1850      17       2320            AA       614  N5BYAA                207
  CRSElapsedTime AirTime ArrDelay DepDelay Origin Dest Distance TaxiIn TaxiOut CarrierDelay WeatherDelay NASDelay SecurityDelay
1            132      99       28       48    DTW  SAV      706      4       9           28            0        0             0
2            159     130       46       -4    MSP  IDA      939      5      74            0            0       46             0
3             98      70       43       55    CVG  PHL      507      4      12            0            0        0             0
4             78      42       21       35    CHS  ATL      259     13       9            0            0        0             0
5             69      47       40       14    ATL  CHS      259     20      28           14            0       26             0
6            210     191       57       60    DFW  BOS     1562      5      11           14            0        0             0
  LateAircraftDelay
1                 0
2                 0
3                43
4                21
5                 0
```