Elizabeth McGuckin

Due Date: Thursday, August 30, 2018, by 5 PM

1. Use the airline data stored in this directory:

/depot/statclass/data/dataexpo2009

In the year 2005, find:

1a. the number of flights that occurred on every day of the year, and

/depot/statclass/data/dataexpo2009

```
emcgucki@scholar-fe05:/depot/statclass/data/dataexpo2009 $
emcgucki@scholar-fe05:/depot/statclass/data/dataexpo2009 $ cat 2005.csv | cut -d, -f2,3 | sort | uniq -c
  16477 10,1
  19885 10,10
  19515 10,11
  19701 10,12
  19883 10,13
  19910 10,14
  16628 10,15
  18910 10,16
  19851 10,17
  19495 10,18
  19732 10,19
  18812 10,2
  19902 10,20
  19928 10,21
  16656 10,22
  18922 10,23
  19864 10,24
  19528 10,25
  19739 10,26
  19899 10,27
  19909 10,28
  16620 10,29
```

1b. find the day of the year on which the most flights occur.

August 5, 2005

```
  20962 7,22
  20977 8,3
  20979 8,10
  21006 8,8
  21013 8,1
  21022 8,11
  21022 8,4
  21037 8,12
  21041 8,5
emcgucki@scholar-fe05:/depot/statclass/data/dataexpo2009 $ cat 2005.csv | cut -d, -f2,3 | sort | uniq -c | sort -
n
```

2. Again considering the year 2005, did United or Delta have more flights?

Delta had more flights.

```
emcgucki@scholar-fe05:/depot/statclass/data/dataexpo2009 $
emcgucki@scholar-fe05:/depot/statclass/data/dataexpo2009 $ cat 2005.csv | cut -d, -f9 | sort | uniq -c | sort -n
       1 UniqueCarrier
   44109 TZ
   48183 HA
   53255 F9
  111050 B6
  136492 DH
  158100 AS
  195392 FL
  197853 HP
  296661 CO
  304647 EV
  381337 OH
  403767 XE
  425609 US
  480832 NW
  485918 UA
  517454 OO
  532032 MQ
  658302 DL
  673569 AA
 1036034 WN
emcgucki@scholar-fe05:/depot/statclass/data/dataexpo2009 $
```

3. Consider the June 2017 taxi cab data, which is located in this folder:

/depot/statclass/data/taxi2018

What is the distribution of the number of passengers in the taxi cab rides? In other words, make a list of the number of rides that have 1 passenger; that have 2 passengers; etc.

```
6933189 1
emcgucki@scholar-fe05:/depot/statclass/data/taxi2018 $ cat yellow_tripdata_2017-06.csv | cut -d, -f4 | sort | uniq
-c
       1
     548 0
 6933189 1
 1385066 2
  406162 3
  187979 4
  455753 5
  288220 6
      26 7
      30 8
      20 9
       1 passenger_count
emcgucki@scholar-fe05:/depot/statclass/data/taxi2018 $
```
0 passengers has 548 rides
1 passenger has 6933189 rides
2 passengers have 1385066 rides
3 passengers have 406162 rides
4 passengers have 187979 rides
5 passengers have 455753 rides
6 passengers have 288220 rides
7 passengers have 26 rides

8 passengers have 30 rides
9 passengers have 20 rides