a) Harsha Thulasi
b) I spent about 6 hours on the lab. Had to review some stuff from the lecture but mostly got stuff from my notes in lecture.
c) It took me a while to understand the 3rd question. I'm assuming I did it as expected but beyond that most of it was similar to what we did in class. The third question where I had to read all the files in textcoropora killed my pc. Trying to recover the commands I used took me a while.
d) I did all three parts. For the third (or "g"), I only collected data from "austen*" files. My computer was crashing if I had tried to run mapreduce on all the files in the directory. Each part is named as "first", "second", "third".
e) Output:

```
root@1f2487d9d2c6:/code# hdfs dfs -cat /output/first/part*
2021-04-10 03:18:10,893 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
01	157	157	0	0	0
02	188	188	0	0	0
03	232	232	0	0	0
04	268	267	0	1	0
05	73	2	71	0	0
06	260	76	151	31	2
07	210	30	150	30	0
08	210	30	150	30	0
09	210	30	150	30	0
10	192	28	136	28	0
root@1f2487d9d2c6:/code#
```

f) Output:

```
root@9c35920c5ffd:/code# hdfs dfs -cat /output/airline_delay/part*
2021-04-10 05:03:41,254 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
AA	0.000000	1659.000000	10.695254
AS	0.000000	567.000000	6.978953
B6	0.000000	746.000000	17.283021
DL	0.000000	1223.000000	9.460750
EV	0.000000	1237.000000	10.915249
F9	0.000000	630.000000	10.652893
HA	0.000000	949.000000	4.184370
NK	0.000000	892.000000	18.634635
OO	0.000000	1370.000000	15.167912
UA	0.000000	1107.000000	9.674842
VX	0.000000	381.000000	15.657723
WN	0.000000	640.000000	7.574056
root@9c35920c5ffd:/code#
```

a) I used the arrival time delay as a means of identifying. That's I personally use to determine a flight is delayed or not.

Sqoop command:

```
sqoop export --connect jdbc:mysql://localhost/airline \
        --username training \
        --password training \
        --table On_Time_On_Time_Performance_2016_1 \
        --columns "UniqueCarrier, ArrDelayMinutes" \
```

--export-dir /database/airline/On_Time_On_Time_Performance_2016_1 \
--input-fields-terminated-by "\t"

g) Output:

```
root@323076c2a13d:/code# hdfs dfs -cat /output/second_third/part* | head
2021-04-10 06:28:46,899 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false,
remoteHostTrusted = false
the     12559
to      12017
and     10794
of      10411
a       6711
her     6133
i       6008
was     5557
in      5459
it      5142
cat: Unable to write to output stream.
root@323076c2a13d:/code#
```

h)

i)

j)