

## INF141/ICS121

Rex Rockwell #29198266

Leon Thai #55997869

Tuan Vo #60819735

1. Create a PDF file called quant.pdf that includes answers to those questions and, for each, a description of the process you followed to find those answers. If you used scripts or programs for this part, include them in a zipped folder called part2.

2. How many people are targeted in this data set?

150 people

I used the following commands to arrive at the answer

`echo * | wc (while in the current directory)`

`ls -l | grep ^d | wc -l (while in the same directory)`

3. How many individual data files are we dealing with?

- 517424 files total in all from 150 people
  - used the following command in unix
    - `find <dir> -type f | wc -l`

4. How many messages were sent by these people in total? (Explain how you're interpreting the term "sent.")

126069 files

method used: `find ./*/*sent*/ -type f | wc -l`

I search directories that has the word sent in it: `_sent_mail`, `sent`, and `sent_items`

5. How many messages were sitting in these people's Inboxes in total? (Explain how you're determining what counts as an "inbox.")

44871 files

method used: `find ./*/inbox/ -type f | wc -l`

I search all the inbox folders and counted the files in them.

6. Who are the 10 people with the largest number of data files?

```
99M ./mann-k  
90M ./jones-t  
83M ./shackleton-s  
67M ./taylor-m  
55M ./farmer-d  
54M ./germany-c  
54M ./beck-s  
52M ./symes-k  
47M ./nemec-g  
  
42M ./scott-s
```

method used: du -dh 1 | sort -rn | head -15

I used this command on linux command line

du means disk usage

-d means depth

-h means make the file size human readable