

## Project 05 Answer Key

### Question 1a

```
# Use read.csv() to read the LA AirBnB data into RStudio
LA_airbnb = read.csv("/class/datamine/data/airbnb/united-states/ca/los-angeles
  /2019-07-08/visualisations/listings.csv")
# Use tapply() to calculate the number of reviews each host ID received
reviews_by_host = tapply(LA_airbnb$number_of_reviews, LA_airbnb$host_id, sum)
# Use sort() to order the hosts by their respective review counts
# Use tail() to get the hosts with the largest number of reviews
tail(sort(reviews_by_host))

>>>
31102550 1648958 41881230 125558867 2622454 21013529
      2666      3038      3088      3575      4542      5505
```

### Question 1b

```
# Use tapply() to get the AirBnB review totals for each LA neighborhood
reviews_by_neighborhood = tapply(LA_airbnb$number_of_reviews,
  LA_airbnb$neighbourhood, sum)
# Use sort() to order the neighborhoods by number of reviews
# Use tail() to get the 6 neighborhoods with the most reviews
tail(sort(reviews_by_neighborhood))

>>>
Hollywood Hills      Silver Lake      Downtown      Long Beach
      38729      41625      56661      58224
Hollywood      Venice
      94767      151629
```

### Question 2a

```
# Use read.csv() to read the flights data into RStudio
flights = read.csv("/class/datamine/data/flights/2019.csv")
# Use paste() to create a new column by combining the ORIGIN and DEST columns
flights$myflightpath = paste(flights$ORIGIN, flights$DEST)
# Use head() to check that the new column exists and was created successfully
head(flights$myflightpath)

>>>
[1] "MSP CVG" "MSP CVG" "MSP CVG" "MSP CVG" "MSP CVG" "MSP CVG"
```

### Question 2b

```
# Use tapply() to get the mean departure delay of each flight path
delay_by_path = tapply(flights$DEP_DELAY, flights$myflightpath, mean)
# Use sort() to order the flight paths by mean departure delay
# Use tail() to get the 6 flight paths with the highest mean departure delay
tail(sort(delay_by_path))

>>>
ANC DFW GEG LWS AUS CMH CMH AUS AZA MEM VPS SRQ
158.5333 210.0000 217.0000 218.0000 297.0000 720.0000
```

### Question 3a

```
# Use read.delim() to read in the election data with the '/' separator
elections = read.delim("/class/datamine/data/election/itcont2020.txt", sep="|")
# Use paste() to create a new column by combining the CITY and STATE columns
elections$location = paste(elections$CITY, elections$STATE, sep="-")
# Use tapply() to get the donation totals for each city
donations_by_location = tapply(elections$TRANSACTION_AMT, elections$location,
    sum)
```

### Question 3b

```
# Use tapply() to get the donation totals for each committee
donations_by_committee = tapply(elections$TRANSACTION_AMT, elections$CMTE_ID,
    sum)
# Use sort() to order the committees by donation totals
# Use tail() to get the 6 committees with the highest total donation totals
tail(sort(donations_by_committee))

>>>
C00697441 C00010603 C00000935 C00695585 C00003418 C00618389
16824028 19137654 19226010 23379977 30287380 37197362
```