MATH 216: Assignment 2

Solution Key

Instructions

- please finish these questions by Wed, Feb 26.
- submit the .html file to Canvas
- you are encouraged to work together and ask your peers questions. Each person should submit their own work.
- You may share parts of your code to ask or answer questions on Slack. You should avoid sharing (copying and pasting) the entirety of your answers.
- make sure you include at least one acknowledgement

Loading the data

The following data is salaries at the University of North Carolina (UNC). Load it the data.

```
#To load directly from webpage
UNCdata <- read.csv("http://ryanthornburg.com/wp-content/uploads/2015/05/UNC_Salares_NandO_2015-05-06.cs
UNCdata <- as_tibble(UNCdata)

#OR
#to load locally (from file saved in same folder)
#UNCdata <- read.csv("UNCdata.csv")
```

Take a look at the data. Note what the column names are. Note what each row represents. Note that `fte` means full-time employee.

Problems

Question 1

Return a data frame with columns: `name`, `dept`, `age`, `totalsal`

```
UNCdata %>%
  select(name, dept, age, totalsal)
```

```
## # A tibble: 12,287 x 4
##    name                      dept                          age totalsal
##    <fct>                     <fct>                       <int>    <int>
##  1 AARON, NANCY G            Romance Languages              55    46350
##  2 ABARBANELL, JEFFERY S     Kenan-Flagler Business School  57   173000
##  3 ABARE, BETSY              Institute of Marine Sciences   54    38170
##  4 ABATE, AARON B            Medicine Administration        29    50070
##  5 ABATEMARCO, JODI M        School of Education            35    41696
##  6 ABBOTT-LUNSFORD, SHELBY L Medicine Administration        41    61000
##  7 ABBOTTS, WILLIAM C        Biology                        62    41707
```

```
##  8 ABDOULAYI, SARA M         Carolina Population Center       36    80227
##  9 ABDULLAH, LUBNA           Cys Fibrosis/Pulmonary Res       64    55803
## 10 ABE, PAIGE                Housing Res Education            26    32889
## # ... with 12,277 more rows
```

Question 2

Rename the `fte` column to `fulltime`. Make sure this change is saved (i.e. `data <- ...`).

```
UNCdata <- UNCdata %>%
  rename(fulltime=fte)
```

Question 3

Return the mean salary in the Neurosurgery department?

```
UNCdata %>%
  filter(dept == "Neurosurgery") %>%
  summarize(mean.totalsal = mean(totalsal))
```

```
## # A tibble: 1 x 1
##   mean.totalsal
##           <dbl>
## 1       380058.
```

Return a data frame with employees in the Neurosurgery department making more than $500,000.

```
UNCdata %>%
  filter(dept == "Neurosurgery") %>%
  filter(totalsal >=500000) %>%
  arrange(-totalsal)  #arrange is optional
```

```
## # A tibble: 6 x 14
##   name  campus dept  position exempt2 employed hiredate fulltime status
##   <fct> <fct>  <fct> <fct>    <fct>      <int>    <int>    <dbl> <fct>
## 1 CAMP~ UNC-CH Neur~ Adjunct~ Exempt        12 20140731        1 Fixed~
## 2 CARS~ UNC-CH Neur~ Clinica~ Exempt        12 20090430        1 Fixed~
## 3 JAUF~ UNC-CH Neur~ Clinica~ Exempt        12 20080930        1 Fixed~
## 4 KILP~ UNC-CH Neur~ Clinica~ Exempt        12 20130930        1 Fixed~
## 5 EWEN~ UNC-CH Neur~ DIRECTOR Exempt        12 19970731        1 Conti~
## 6 WADO~ UNC-CH Neur~ Clinica~ Exempt        12 20080930        1 Fixed~
## # ... with 5 more variables: stservyr <int>, statesal <int>,
## #   nonstsal <int>, totalsal <int>, age <int>
```

Question 4

What is the total amount that full time Dermatology employees get paid?

```
UNCdata %>%
  filter(dept=="Dermatology") %>%
  filter(fulltime==1) %>%
  summarize(salary.total=sum(totalsal))
```

```
## # A tibble: 1 x 1
##   salary.total
##          <int>
## 1      5272098
```

Question 5

How many departments have at least 10 employees?

```
UNCdata %>%
  group_by(dept) %>%
  summarize(count=n()) %>%
  filter(count >=10) %>%
  nrow()
```

```
## [1] 194
```

Question 6

Create a data frame called `radio_dept` whose rows are the employees from the Radiology department.

- include only the following columns: `name`, `position`, `age`, `nonstsal`, `totalsal`.
- order the employees by `salary`

```
radio_dept <- UNCdata %>%
  filter(dept == "Radiology") %>%
  select(name, position, age, nonstsal, totalsal) %>%
  arrange(totalsal)

radio_dept
```

```
## # A tibble: 88 x 5
##    name               position                        age nonstsal totalsal
##    <fct>              <fct>                         <int>    <int>    <int>
##  1 HOOTS, TIFFANY N   Social/Clinical Research Ass~    31    36360    36360
##  2 FISCHER, MICHELLE~ Admin. Support Associate         25    37142    37142
##  3 BIRDSONG, LAURIE B Public Communications Specia~    40    37681    37681
##  4 PENDER, JENNIFER L Accounting Technician            39    37690    37690
##  5 BARBAL, ISABEL     Admin. Support Associate         57    40061    40061
##  6 MELVILLE, WILMA C  Administrative Secretary II      58    41789    41789
##  7 HARTMAN, TERRY S   Social/Clinical Research Ass~    26    42168    42168
##  8 BOOMHOWER, JEREMY~ Admin. Support Associate         38    42593    42593
##  9 CARVER, VIRGINIA B Admin. Support Associate         39    42593    42593
## 10 HAUSER, JASON M    Admin. Support Associate         41    42593    42593
## # ... with 78 more rows
```

Question 7

Create a data frame called `dept_summary` whose rows are the departments and whose columns are: department size, mean department salary, median department salary, and maximum salary (using `totalsal` for salary).

```
dept_summary <-UNCdata %>%
  group_by(dept) %>%
  summarize(dept.size = n(),
            mean.dept.sal = mean(totalsal),
            med.dept.sal = median(totalsal),
            max.dept.sal = max(totalsal))

dept_summary
```

```
## # A tibble: 304 x 5
##    dept                    dept.size mean.dept.sal med.dept.sal max.dept.sal
##    <fct>                       <int>         <dbl>        <dbl>        <int>
```

```
##  1 Acad Sup Prog Student~        15        55798.       50600        115000
##  2 Academic Advising             42        49985.       45000        109625
##  3 Accounting Services           17        57417.       59342        103306
##  4 Ackland Art Museum            19        51543.       41000        140050
##  5 Admissions                    46        57487.       49000        195700
##  6 African Studies Center         2        35970        35970         43475
##  7 African, Afri-Amer & ~        23        65170.       68000        135608
##  8 AHEC Support-Comm Med~        26        69789.       64533        135193
##  9 Airport                        1        47351        47351         47351
## 10 Alcohol Studies Center        16        49232.       49180.        84685
## # ... with 294 more rows
```

Order the departments by highest mean salary and print the 10 highest paid departments, on average.

```
dept_summary %>%
  select(dept, mean.dept.sal) %>%
  arrange(-mean.dept.sal) %>%
  head(10)
```

```
## # A tibble: 10 x 2
##    dept                     mean.dept.sal
##    <fct>                            <dbl>
##  1 Neurosurgery                   380058.
##  2 Provost                        273790
##  3 Urology                        216291.
##  4 Orthopaedics                   216205.
##  5 Surgery                        201917.
##  6 Anesthesiology                 187177.
##  7 Radiation Oncology             183045.
##  8 Carolina Counts                182160
##  9 Radiology                      172053.
## 10 Office of the Chancellor       164747.
```

Order the departments by highest median salary and print the 10 highest paid departments, on average.

```
dept_summary %>%
  select(dept, med.dept.sal) %>%
  arrange(-med.dept.sal) %>%
  head(10)
```

```
## # A tibble: 10 x 2
##    dept                     med.dept.sal
##    <fct>                           <dbl>
##  1 Neurosurgery                   395550
##  2 Provost                        240080
##  3 Orthopaedics                   240000
##  4 Urology                        237500
##  5 Anesthesiology                 222645
##  6 Carolina Counts                182160
##  7 Radiation Oncology             180000
##  8 Surgery                        176083
##  9 University Ombuds Office       157127
## 10 Ath Basketball Office          150000
```

Why do these lists differ? If you were asked for the top 10 best paid departments at UNC which summary would you choose and why?

Median is calculated as the department member who makes a salary exactly in the middle, meaning half

their colleagues make more than they do and half their colleagues make less than they do. Mean is calculated differently - it takes the sum of the salaries and divides by the total number of employees. Means are much more effected by outlying (very high or very low) salaries.

For example, the Office of the Chancellor appears on the top ten departments with highest mean salary but does not appear on the list of top ten departments with highest median salaries. This is because the Chancellor has a very high salary ($520,000), but the remaining salaries in that department are much less.

Question 8

Make a list of all the department names and sort this list alphabetically. What is the 42nd department in this list?

```
dept_summary %>%
  select(dept) %>%
  arrange(dept) %>%
  slice(42)
```

```
## # A tibble: 1 x 1
##   dept
##   <fct>
## 1 Ath Soccer
```

Acknowledgements

Use this space to acknowledge anyone who has helped you with this lab. This could be a peer who helped you when you got stuck. This could be the peer tutor. This could be your family or a friend for their support. **You must include at least one acknowledgement**.