Questions? Contact Sarah Hutchins, shutchins@spotlightpa.org

Practice

Your turn! How many employees in our data? (You may already know the answer to this, but write some code anyway!)

ANSWER: 1200

```
salaries %>%
  summarise(count = n())
```

Your turn! Who made the most in total compensation? Who made the least? (Hint: use arrange() to sort your data)

Answer: Scott McKnight made most; Susan Yoon made the least.

Hint: Show them how to select specific columns if they want.

```
salaries %>%
  select(last_name, first_name, total_comp) %>%
  arrange(desc(total_comp))
```

Your turn! Who made the most in overtime/oncall pay?

Answer: Jeffrey Rodgers made most; Sarah Moore made the least

```
salaries %>%
  select(last_name, first_name, department, overtime_oncall) %>%
  arrange(desc(overtime_oncall))
```

Your turn! What's the average and median salary for 2021? Hourly rate? (Note: both of these have NAs, so code accordingly)

Avg. Hourly = 17.22 Median Hourly = 15.18

Avg Salary = 55,339 Median Salary = 56,325

```
salaries %>%
  summarise(avg_hourly = mean(hourly_rate, na.rm=T),
            median_hourly = median(hourly_rate, na.rm=T))
```

Getting to know your data

Your turn! Try using the count() function on department. How clean are the department names?

Answer: Not bad until we get to Utilities!

```
salaries %>%
  count(department) %>%
  arrange(department)
```

Your turn! Use the filter() to look at the rows for Emily Herr. What can we learn about her work. Does it make sense that she's in here twice, or is this potentially an error in the data?

```
salaries %>%
  filter(last_name == "Herr" & first_name == "Emily")
```

Asking questions

How many people work for the police department?

Answer: 172

```
salaries %>%
  filter(department == "Police")
```

What's the average total compensation for a police employee?

Answer: 51260

```
salaries %>%
  filter(department == "Police") %>%
```

```
  summarise(avg_pay = mean(total_comp))
```

Your turn! Calculate the average compensation for each job title within the Police department:

Answer: Chief, Deputy Chief, Captain have highest avg pay

```
salaries %>%
  filter(department == "Police") %>%
  group_by(job_title) %>%
  summarise(avg_pay = mean(total_comp))
```

Your turn! Let's find the same calculations for the job titles. For each job title, calculate the following:

- Average compensation
- Total compensation
- Number of people with that job title

Arrange your results by the job title that has the highest average compensation.

```
salaries %>%
  group_by(job_title) %>%
  summarise(avg_comp = mean(total_comp),
            total_comp = sum(total_comp),
            num_employees = n()) %>%
  arrange(desc(avg_comp))
```

Your turn! Which department pays the highest average salary to people with director in their job title?

Answer: Planning & Transportation, 103446.36

```
directors %>%
  group_by(department) %>%
  summarise(avg_salary = mean(salary_2021)) %>%
  arrange(desc(avg_salary))
```

Your turn! Use filter and grepl to find all the people who work in the various Utilities departments. (If you need to refresh your memory, click on the word "salaries" in your Environment (upper right).) Once you've successfully run this code, store these employees to their own data frame called utilities.

Answer: Should end up w/ 217 employees/rows in new table.

```
utilities <- salaries %>% filter(grepl("Utilities", department))
```

Your turn! Which job in the various Utilities department pays the best? (This question is intentionally vague! Think about the various calculations you can do and pick one – or multiple – to try to come up with a conclusion.)

Answer: Director has highest avg comp, but there are not many of them.

```
utilities %>%
  group_by(job_title) %>%
  summarise(avg_comp = mean(total_comp),
            median_comp = median(total_comp),
            avg_salary = mean(salary_2021),
            median_salary = median(salary_2021),
            avg_overtime = mean(overtime_oncall, na.rm=T)) %>%
  arrange(desc(avg_comp))
```

Extra practice!

1. What do people with the word 'specialist' in their job title make, on average?

   Answer: 19555.46

```
salaries %>%
  filter(grepl("Specialist", job_title)) %>%
  summarise(avg_pay = mean(total_comp))
```

2. What do interns make?

Answer: Avg total compensation of 4339.417

```
salaries %>%
```

```
  filter(grepl("Intern", job_title)) %>%
  summarise(avg_pay = mean(total_comp))
```

3. Which department paid out the most in overtime/on-call pay?

Answer: Police with 746006.10 in total_ot

```
salaries %>%
  group_by(department) %>%
  summarise(total_ot = sum(overtime_oncall, na.rm=T)) %>%
  arrange(desc(total_ot))
```

4. Which department has the most employees paid hourly?

Answer: Parks & Rec, 325 people

```
salaries %>%
  filter(hourly_rate > 0) %>%
  group_by(department) %>%
  summarise(num_employees = n()) %>%
  arrange(desc(num_employees))
```

5. For police employees, find the percent of their total compensation comes from overtime for each employee.

Do this in two steps: First, create a data frame called police of just employees who work for the police department. If you do this correctly, you will see police show up in your Environment sidebar. Then, using this new police dataframe, you will use mutate() to add a column and do a percent of total calculation.

```
police <- salaries %>% filter(department == "Police")

police %>%
  mutate(pct_ot = overtime_oncall/total_comp) %>%
  arrange(desc(pct_ot))
```