Retrieval Practice and Learning

By Ivan Neto

What is the most effective way to learn a subject? Many students focus exclusively on the *encoding* process---that is, how to get the knowledge into memory in the first place. For example, taking notes is an activity for encoding knowledge.

Retrieval, on the other hand, is the process of reconstructing that knowledge from memory. Karpicke and Blunt (2011) demonstrated that *retrieval* is more effective for learning than activites designed to promote effective encoding. They conducted an experiment in which subjects had to learn about sea otters by reading a passage. Subjects were randomly assigned to one of two conditions: some were instructed to create a concept map as they read the passage, while others were instructed to practice retrieval (i.e., read the passage, recall as much as they could, read the text again, and recall again). The two main measurements they recorded were:

1. each subject's score on a follow-up learning test one week later
2. each subject's *prediction* of how well they would do on that test

In this lab, you will analyze data from a *replication* of Karpicke and Blunt's experiment, conducted by Buttrick *et al.*

- The data file is : data.csv.
- The codebook (explaining what the variables mean) is : codebook.csv.

In [ ]:
```
# READ IN THE DATA SET HERE

import pandas as pd
import matplotlib.pyplot as plt
from IPython.display import display
import seaborn as sns
import numpy as np
import math

pd.options.display.max_columns = 35

df = pd.read_csv("data.csv")

df.head(10)
```

Out[ ]:

	ID	Age	Gender	Date.P1	Date.P2	Condition	IC.1	IC.2	Comp.1	Comp.2	PR.1	PR.2	MC
0	KB1	18	Female	11/21/16	11/28/16	Concept	1	1	1	1	1.0	0.80	0.7
1	KB2	18	Male	11/21/16	11/28/16	Concept	1	1	1	1	0.0	0.70	0.4
2	KB3	18	Male	11/21/16	11/28/16	Concept	1	1	1	1	0.0	0.70	0.4
3	KB4	19	Female	11/21/16	11/28/16	Concept	1	1	1	1	0.0	0.70	0.3
4	KB5	19	Female	11/22/16	11/29/16	Concept	1	1	1	1	0.0	0.95	0.6
5	KB6	19	Male	11/22/16	11/29/16	Concept	1	1	1	1	1.0	0.60	0.8
6	KB7	18	Male	11/22/16	12/6/16	Concept	1	1	1	1	1.0	0.60	0.7
7	KB8	20	Male	11/22/16	11/29/16	Concept	1	1	1	1	1.0	0.50	0.3
8	KB9	20	Male	11/22/16	11/28/16	Concept	1	1	1	1	0.0	0.60	0.6
9	KB10	20	Female	11/29/16	12/6/16	Concept	1	1	1	1	1.0	0.90	0.8

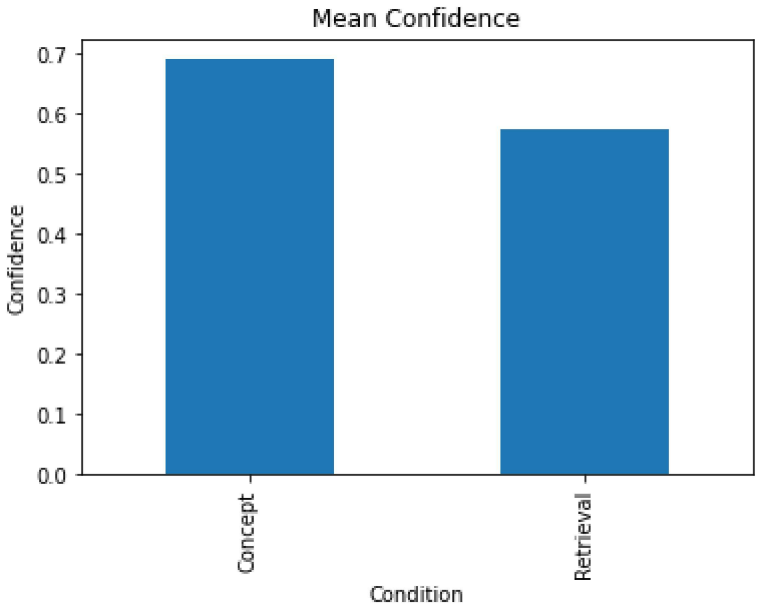
Question 1

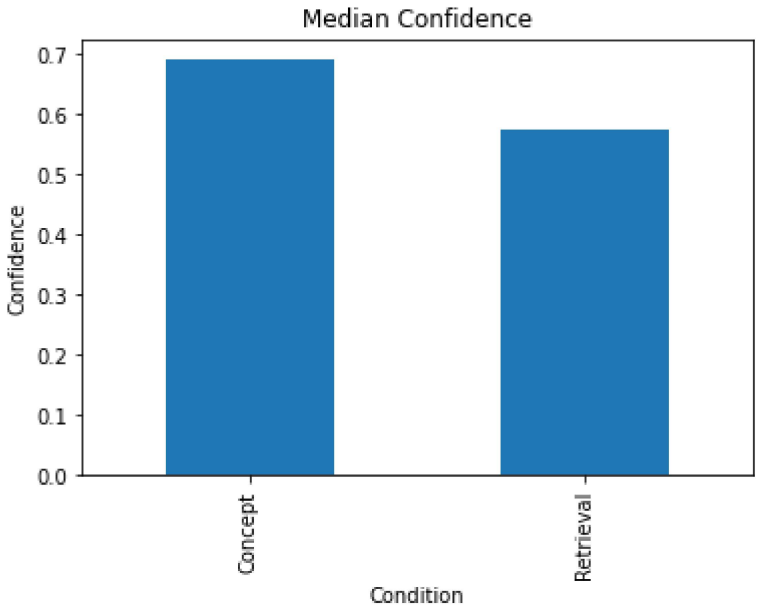
Which group felt like they learned more: the subjects who made concept maps or the ones who practiced retrieval? (Or are they about the same?) Make an appropriate visualization and explain what you see.

Hint: Use the variable `PR.2`, which contains the participants' predictions of how well they would do on a test one week later.

In [ ]:

```
# YOUR CODE HERE

# By mean
df.groupby("Condition")["PR.2"].mean().plot(kind="bar")
plt.ylabel("Confidence")
plt.title("Mean Confidence")
plt.show()

# By median
df.groupby("Condition")["PR.2"].mean().plot(kind="bar")
plt.ylabel("Confidence")
plt.title("Median Confidence")
plt.show()
```





YOUR EXPLANATION HERE

- It looks like the subjects who made concept maps felt like they would perform better in the test on the upcoming week.

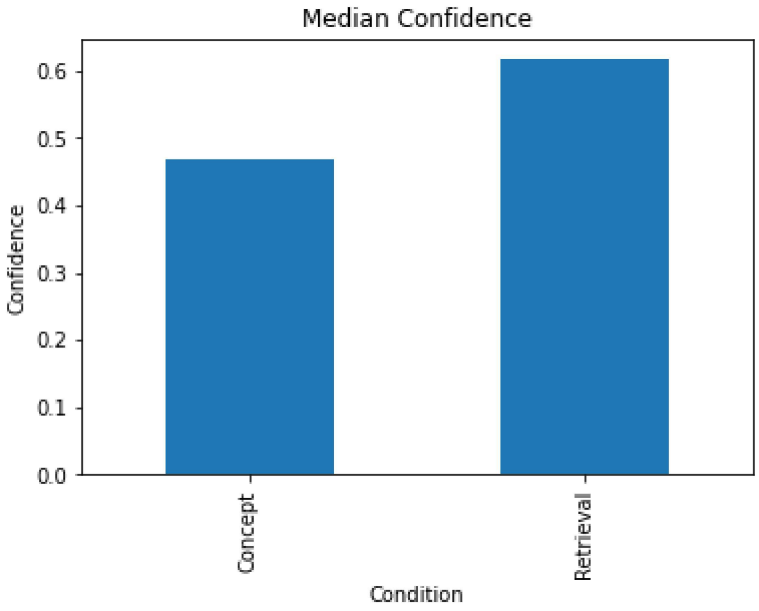
Question 2

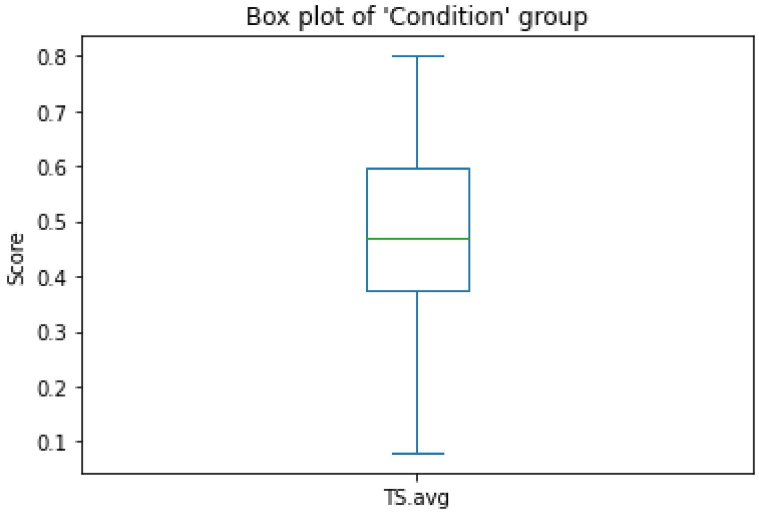
Which group actually did better on the follow-up learning test one week later? Make an appropriate visualization and explain what you see.

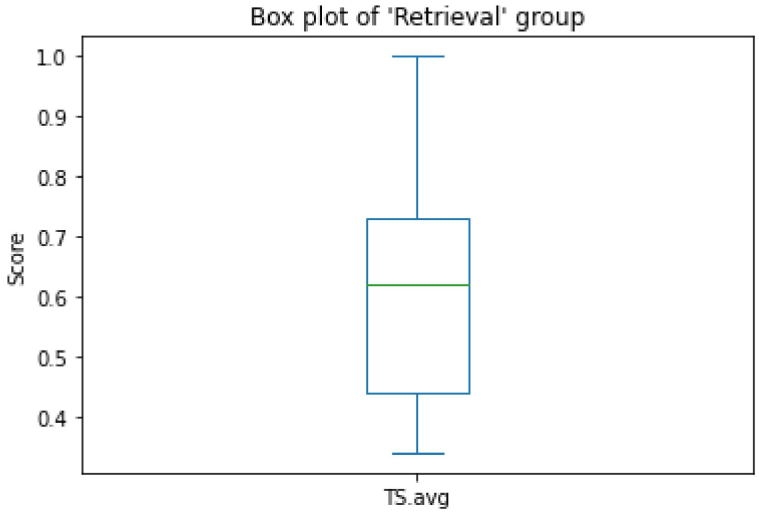
Hint: Don't ask which variable you should use. That is for you to figure out. Read the codebook carefully (consulting the original paper, if necessary), make an informed decision, and explain your choice.

In [ ]:
```
# YOUR CODE HERE
```

```
# By Mean
df.groupby("Condition")["TS.avg"].mean().plot(kind="bar")
plt.ylabel("Confidence")
plt.title("Mean Confidence")
plt.show()

# By Median
df.groupby("Condition")["TS.avg"].mean().plot(kind="bar")
plt.ylabel("Confidence")
plt.title("Median Confidence")
plt.show()

# Entire distribution of TS.avg - "Concept" group
df.loc[df["Condition"] == "Concept"]["TS.avg"].plot(kind="box")
plt.ylabel("Score")
plt.title("Box plot of 'Condition' group")
plt.show()

# Entire distribution of TS.avg - "Concept" group
df.loc[df["Condition"] == "Retrieval"]["TS.avg"].plot(kind="box")
plt.ylabel("Score")
plt.title("Box plot of 'Retrieval' group")
plt.show()
```







YOUR EXPLANATION HERE

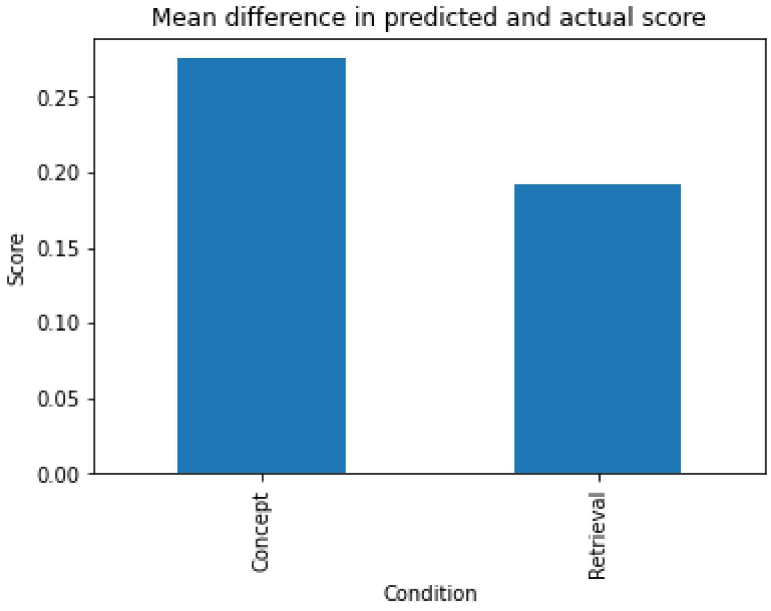
It looks like the group who retrieved the information as they were learning it performed better, since both mean and medians in the "Retrieval" group are higher than in the "Concept" group. Further, the Retrieval group shows a higher mean in the box splots.

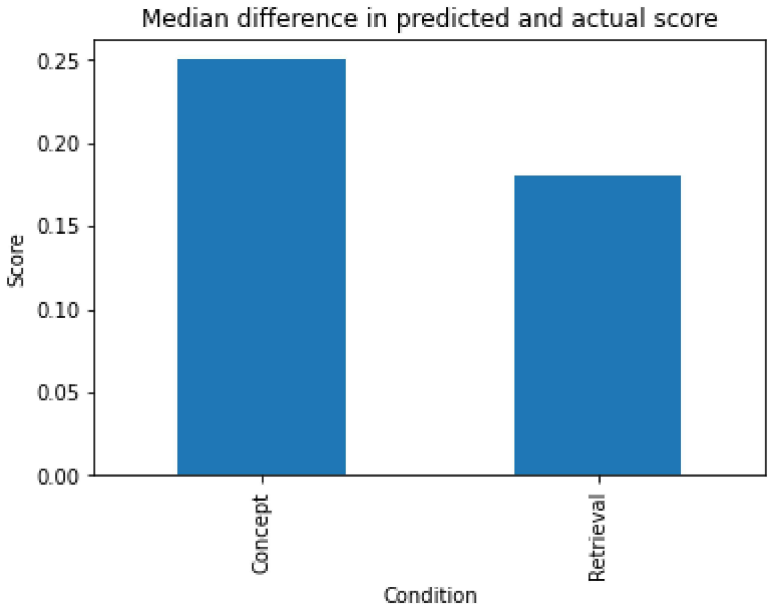
Question 3

How good were subjects at predicting how well they would do on the follow-up learning test? Calculate a measure of how well subjects predicted their performance and interpret the value in context. (Optionally, you may want to include a visualization as well.)

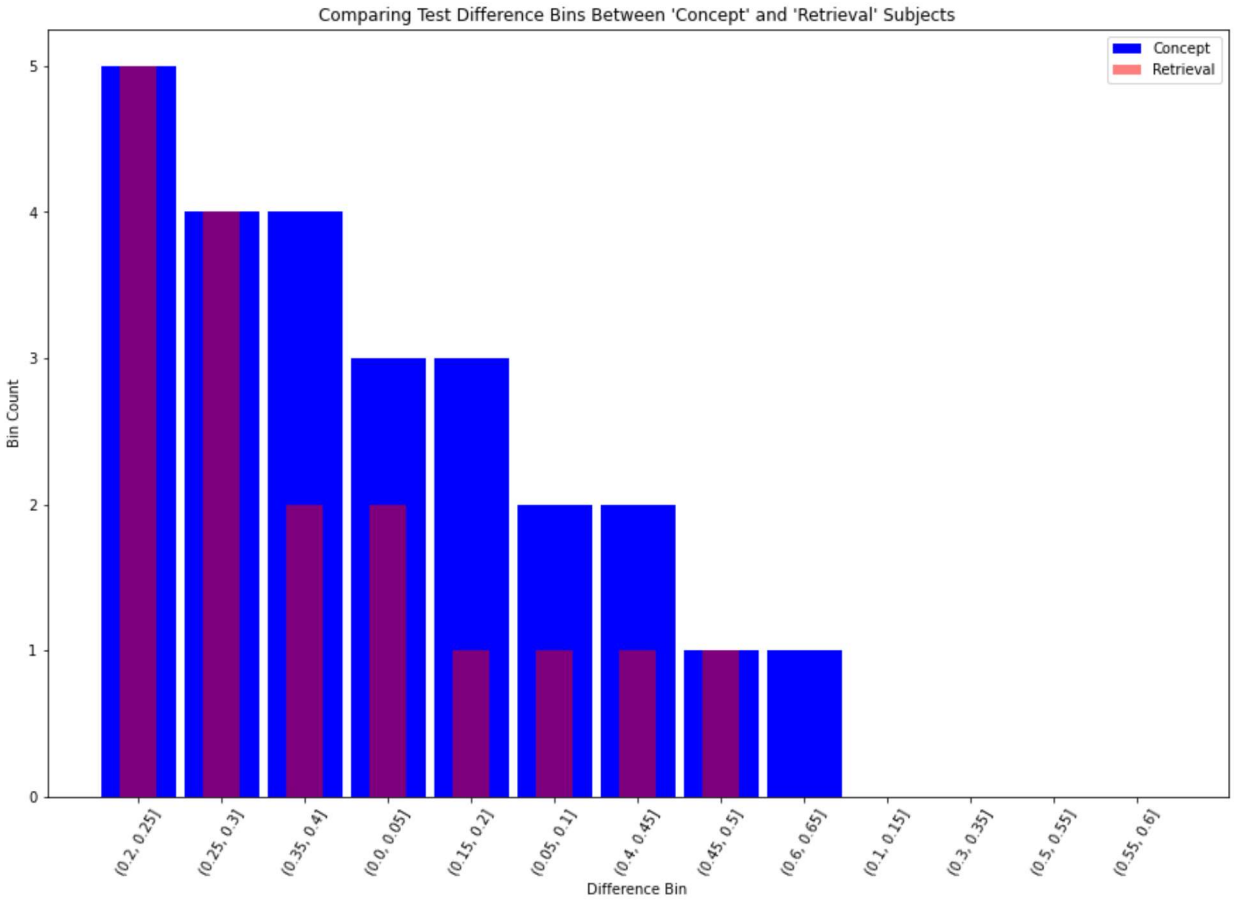
In [ ]:
```
# YOUR CODE HERE

# Create measure of difference between PR.2 and TS.2
df["TDIF"] = (df["PR.2"] - df["TS.avg"]).abs()

# By mean
mean_diff = df.groupby("Condition")["TDIF"].mean()
mean_diff.plot(kind="bar")
plt.title("Mean difference in predicted and actual score")
plt.ylabel("Score")
plt.show()

# By median
median_diff = df.groupby("Condition")["TDIF"].median()
median_diff.plot(kind="bar")
plt.title("Median difference in predicted and actual score")
plt.ylabel("Score")
plt.show()

print("Analysis of Distribution of Differences")
print(" Concept")
print(f"    - Max: {df.loc[df['Condition'] == 'Concept']['TDIF'].max()}")
print(f"    - Min: {df.loc[df['Condition'] == 'Concept']['TDIF'].min()}")
print(f"    - Range: {df.loc[df['Condition'] == 'Concept']['TDIF'].max() - df.loc[df['
print(" Retrieval")
print(f"    - Max: {df.loc[df['Condition'] == 'Retrieval']['TDIF'].max()}")
print(f"    - Min: {df.loc[df['Condition'] == 'Retrieval']['TDIF'].min()}")
print(f"    - Range: {df.loc[df['Condition'] == 'Retrieval']['TDIF'].max() - df.loc[df

print("\nDistributions:")

# Ranges used for visualization
diff_ranges = np.arange(0, 0.7, 0.05)

df["TDIF_BIN"] = pd.cut(x=df["TDIF"], bins=diff_ranges)

concept_counts = df.loc[df["Condition"] == "Concept"]["TDIF_BIN"].value_counts()
retrieval_counts = df.loc[df["Condition"] == "Retrieval"]["TDIF_BIN"].value_counts()

diff_bins = concept_counts.index

plt.figure(figsize=(15, 10))
width = 0.90
plt.bar(x=[i for i in range(0, len(diff_bins))], height=concept_counts, color="blue",

plt.bar(x=[i for i in range(0, len(diff_bins))], height=retrieval_counts, color="red",
plt.xlabel("Difference Bin")
plt.ylabel("Bin Count")
plt.title("Comparing Test Difference Bins Between 'Concept' and 'Retrieval' Subjects")
plt.xticks([i for i in range(0, len(diff_bins))], [str(i) for i in diff_bins], rotatio
```

```
plt.legend()
plt.show()
```






```
Analysis of Distribution of Differences
 Concept
    - Max: 0.6799999999999999
    - Min: 0.010000000000000009
    - Range: 0.6699999999999999
 Retrieval
    - Max: 0.43999999999999995
    - Min: 0.020000000000000018
    - Range: 0.41999999999999994

Distributions:
```



YOUR EXPLANATION HERE

- I chose to use the difference between the predicted score, and the actual score for each group to compare which group performed best at predicting their score. I compared both the mean and median of this new "difference score." The results show that the 'Retrieval' group had a better performance in their score prediction. This is because the difference scores for the 'Concept' group seem to be much higher, indicating that they were worse at predicting their score.
- Further, I display the distribution of differences and compare both of the groups. The red bars are the 'Retrieval' group, which we can see has a generally lower distribution, with every red bar being less than or equal that of the blue bars. The blue bars are generally higher, meaning that more values fell within the bins for blue.

Question 4

This was a completely randomized experiment. This means that the condition that each subject was assigned to should be independent of their gender, age, and any other subject characteristics. Does that seem to be true in this case? Calculate a summary measure and/or make a visualization, and explain what you see.

In [ ]:
```
# YOUR CODE HERE
```

```
# Print the counts
print("Counts:")
print(f"    - Males: {df['Gender'].value_counts()['Male']}")
print(f"    - Females: {df['Gender'].value_counts()['Female']}")
print(f"    - Total: {df['Gender'].count()}\n")

# Contingency Table
print("Contingency Table:")
cont_tab = pd.crosstab(df["Condition"], df["Gender"], margins=True)
display(cont_tab)

# Joint Distribution Table
print("Joint Distribution Table:")
joint_tab = pd.crosstab(df["Condition"], df["Gender"], normalize=True)
display(pd.crosstab(df["Condition"], df["Gender"], normalize=True, margins=True))

# Concept/Retrieval given Gender
gender = joint_tab.sum(axis=0)
condition_given_gender = joint_tab.divide(gender, axis=1)
condition_given_gender.T.plot.bar(stacked=True)
plt.title("Condition Given Gender")
plt.legend(loc="best")
plt.show()

# Gender given Concept/Retrieval
condition = joint_tab.sum(axis=1)
gender_given_condition = joint_tab.divide(condition, axis=0)
gender_given_condition.plot.bar(stacked=True)
plt.title("Gender Given Condition")
plt.legend(loc="best")
plt.show()

# Generate heatmap
sns.heatmap(joint_tab)
plt.show()

# Plot age distributions between Concept and Retrieval subjects
pd.crosstab(df["Condition"], df["Age"]).plot(kind="bar")
```

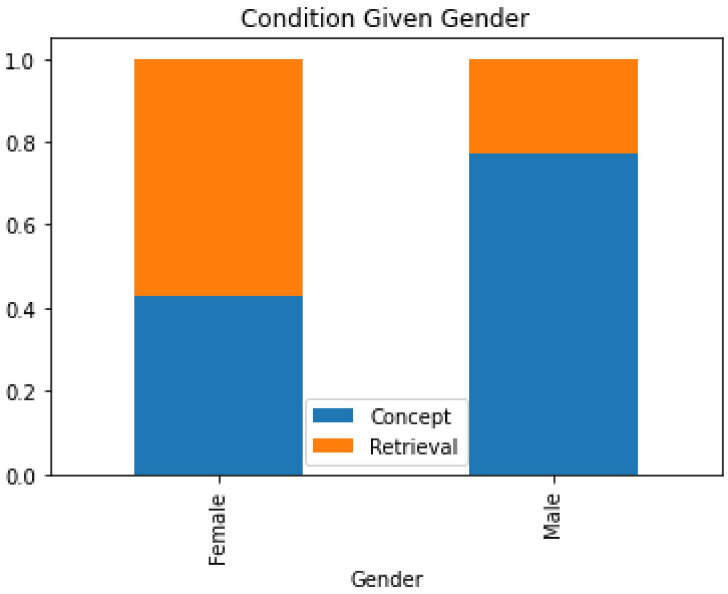
```
Counts:
    - Males: 22
    - Females: 21
    - Total: 43
```

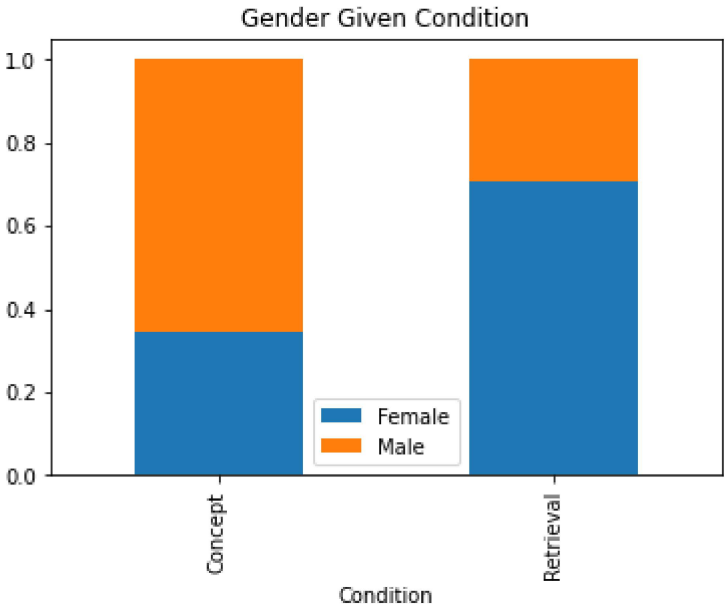
Contingency Table:

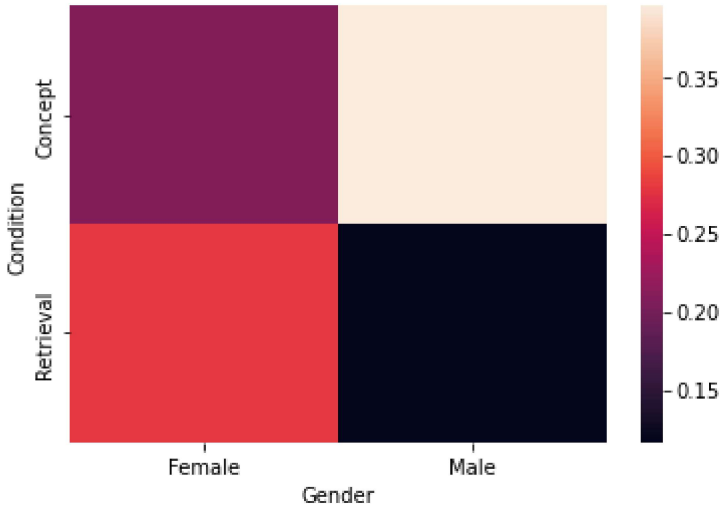
Gender	Female	Male	All
Condition			
Concept	9	17	26
Retrieval	12	5	17
All	21	22	43

Joint Distribution Table:

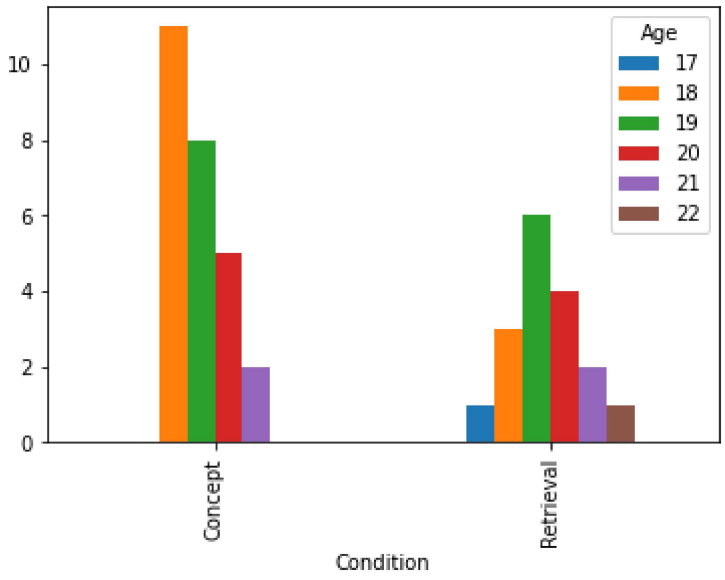
Gender	Female	Male	All
Condition			
Concept	0.209302	0.395349	0.604651
Retrieval	0.279070	0.116279	0.395349
All	0.488372	0.511628	1.000000







Out[ ]: <AxesSubplot:xlabel='Condition'>



YOUR EXPLANATION HERE

- It seems like the distribution of the categories are not quite equal for males and females. Observing the number of males and females, they are approximately equal. Thus, discrepancy between the sample sizes of males and females is not an issue.
- In the first bar graph, we can observe that the 'Concept' group given Male subjects is much larger than the 'Concept' group given Female subjects. Further, both the heat map and second bar graph seem to indicate that out of the 'Concept' group, more subjects are male than female.
- Therefore, it is not true that we observe similar distributions of 'Concept' and 'Retrieval' methods for both males and females. This may be because certain types of methods are preferred by a gender more than the other methods. For example, males may prefer the 'Concept' map method, rather than the 'Retrieval' method. Thus, males will dominate in that category when it comes to representation.

Submission Instructions

Once you are finished, follow these steps:

1. Restart the kernel and re-run this notebook from beginning to end by going to `Kernel > Restart Kernel and Run All Cells`.
2. If this process stops halfway through, that means there was an error. Correct the error and repeat Step 1 until the notebook runs from beginning to end.
3. Double check that there is a number next to each code cell and that these numbers are in order.

Then, submit your lab as follows:

1. Go to `File > Export Notebook As > PDF`.
2. Double check that the entire notebook, from beginning to end, is in this PDF file. (If the notebook is cut off, try first exporting the notebook to HTML and printing to PDF.)
3. Upload the PDF to Gradescope and Notebook (ipynb) to iLearn.