

3C-Solution

October 25, 2021

1 Retrieval Practice and Learning

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 activities 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.

```
[1]: # READ IN THE DATA SET HERE
import pandas as pd
df_data = pd.read_csv('data.csv')
```

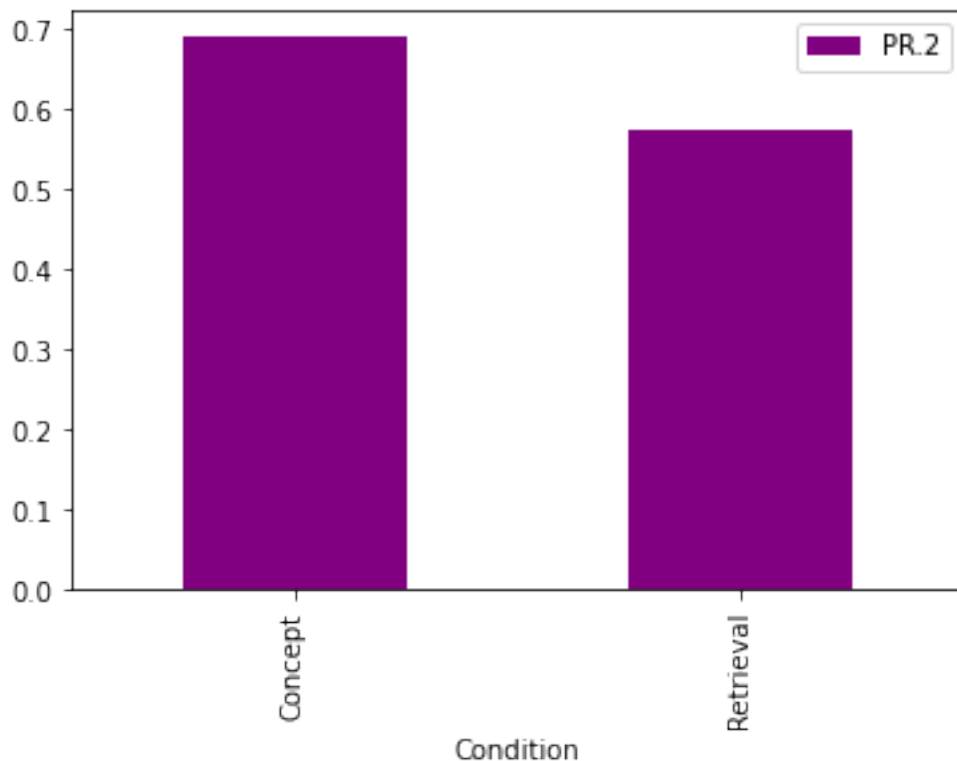
2 Question 1

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.

```
[2]: # YOUR CODE HERE
df_expected = df_data.pivot_table(index="Condition", values="PR.2",
    ↪aggfunc="mean")
df_expected.plot.bar(color = "purple")
```

```
[2]: <AxesSubplot:xlabel='Condition'>
```



YOUR EXPLANATION HERE As seen above, the participant in the concept learning have predicted a higher score as compared to the retrieval learning.

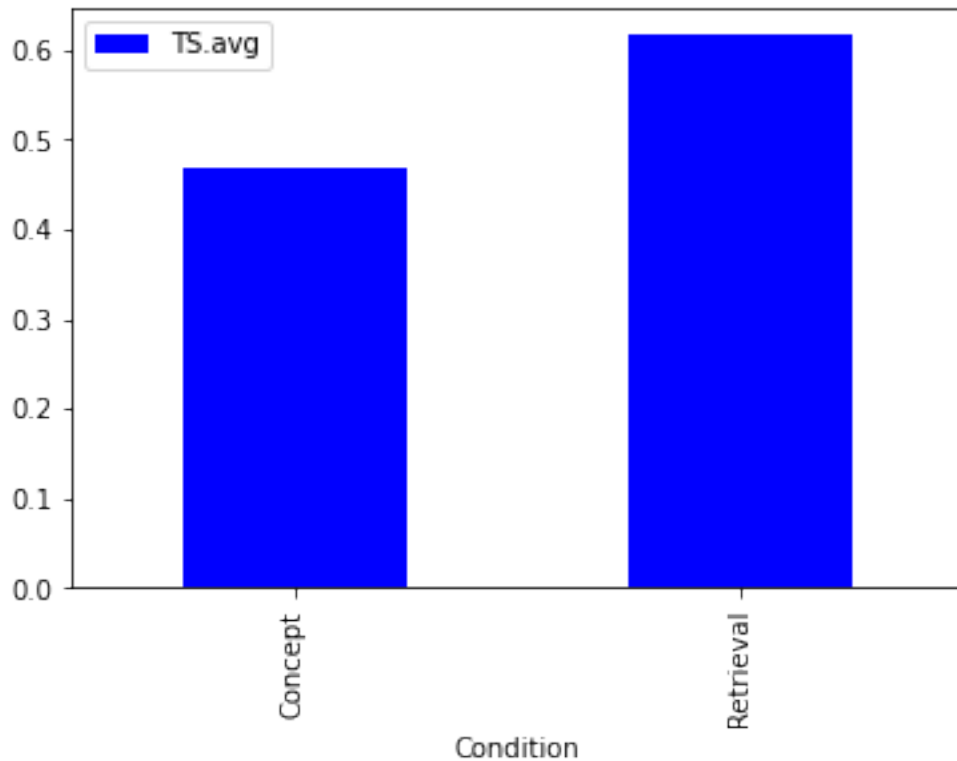
3 Question 2

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.

```
[3]: # YOUR CODE HERE
df_actual = df_data.pivot_table(index="Condition", values="TS.avg",
    ↳aggfunc="mean")
df_actual.plot.bar(color = "blue")
```

```
[3]: <AxesSubplot:xlabel='Condition'>
```



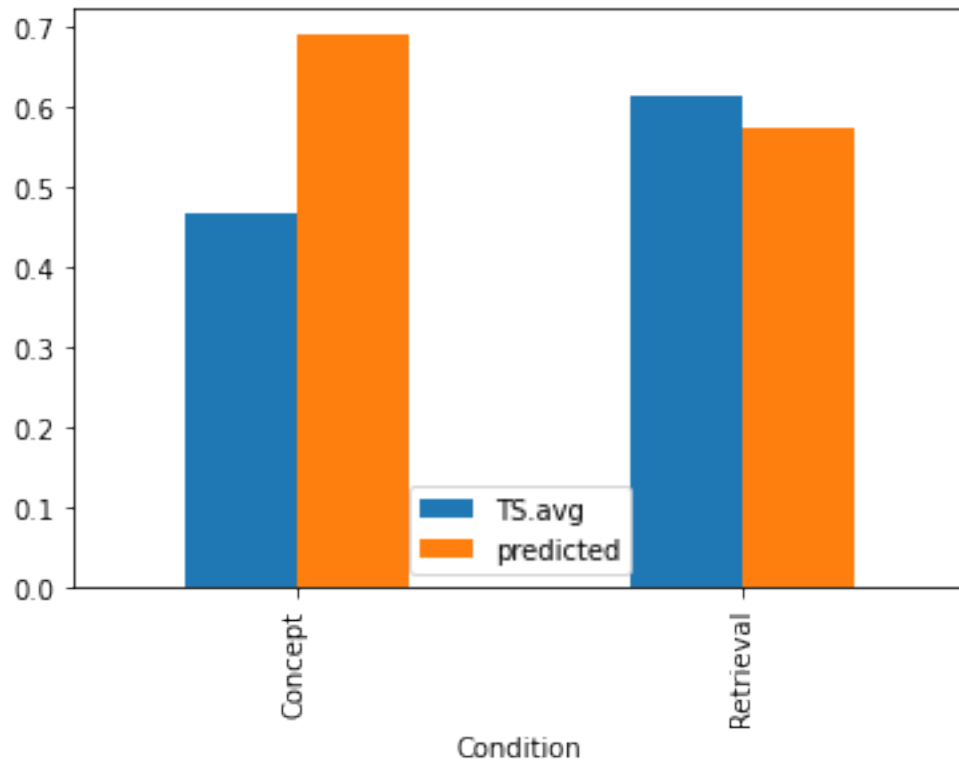
YOUR EXPLANATION HERE As seen above, the participant of concept learning scored lower in the actual test on average in comparison with the participants in the retrieval learning.

4 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.)

```
[4]: # YOUR CODE HERE
df_actual['predicted'] = df_expected['PR.2']
df_actual.plot.bar()
```

```
[4]: <AxesSubplot:xlabel='Condition'>
```



YOUR EXPLANATION HERE Concept learning participants over-estimated their score as opposed to the retrieval learning which under-estimated their scores.

5 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.

```
[5]: # YOUR CODE HERE
df_age = pd.crosstab(df_data.Condition, df_data.Age)
df_avg_age = df_data.pivot_table(index="Condition", values="Age",
    ↪aggfunc="mean")
df_gender = pd.crosstab(df_data.Condition, df_data.Gender)

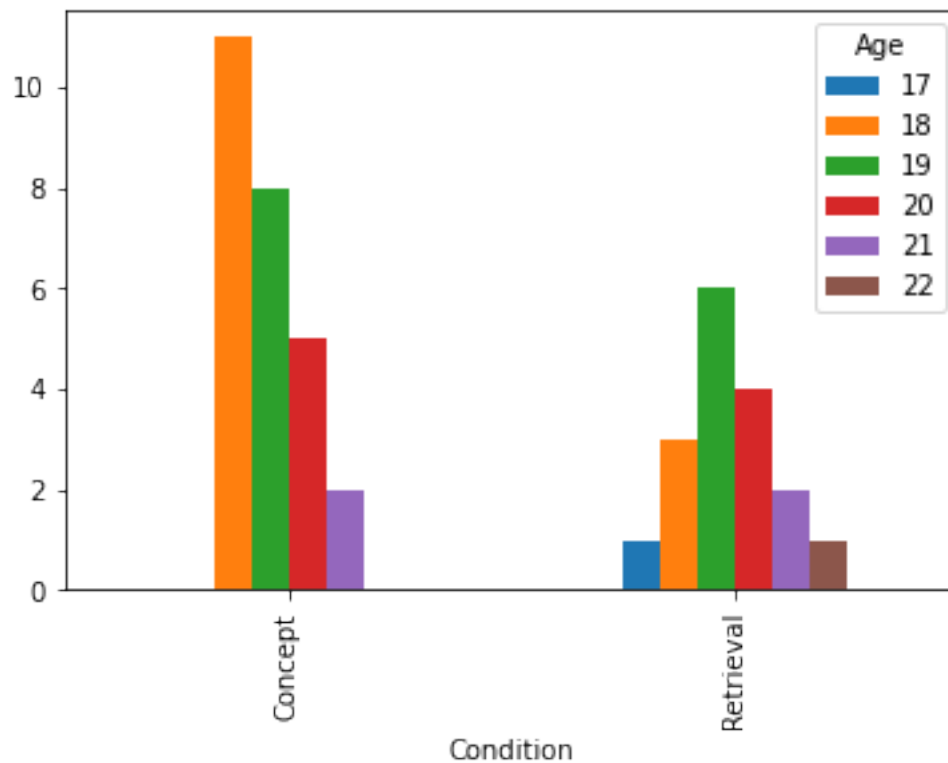
# Plots
print(df_age)
print(df_avg_age)
print(df_gender)
df_age.plot.bar()
df_avg_age.plot.bar()
```

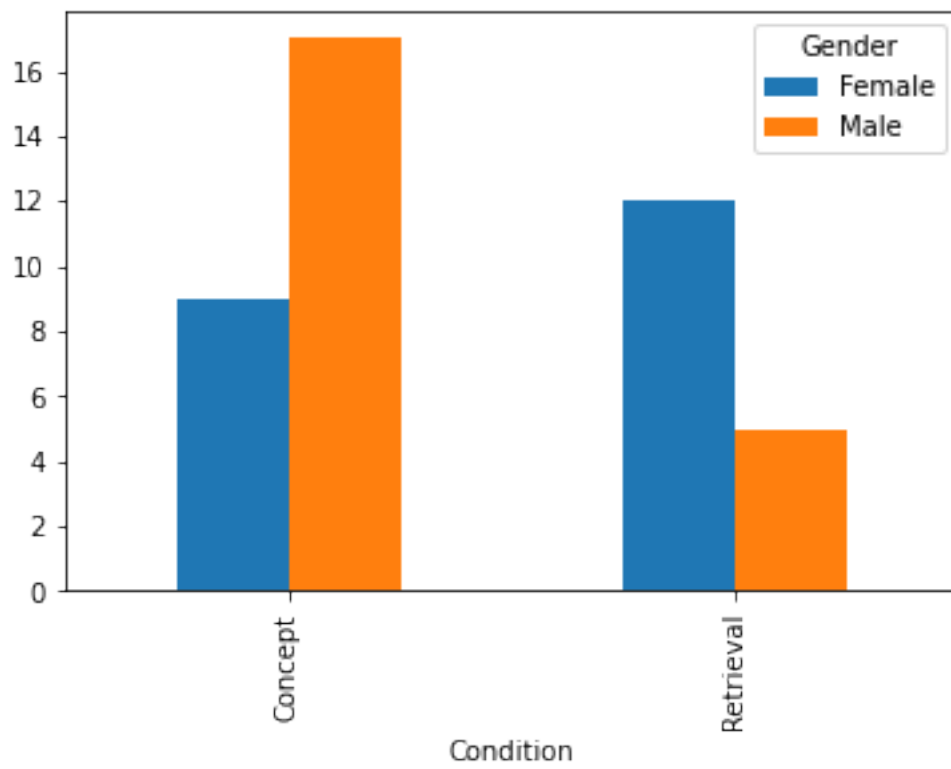
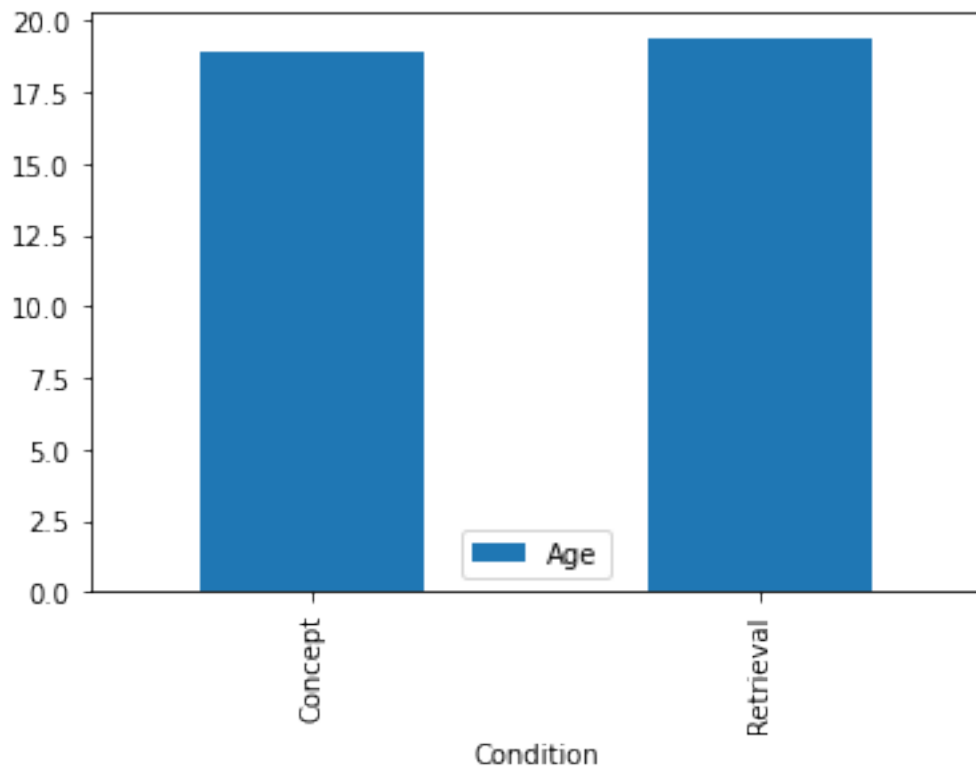
```
df_gender.plot.bar()
```

Age	17	18	19	20	21	22
Condition						
Concept	0	11	8	5	2	0
Retrieval	1	3	6	4	2	1

	Age	
Condition		
Concept	18.923077	
Retrieval	19.352941	
Gender	Female	Male
Condition		
Concept	9	17
Retrieval	12	5

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





YOUR EXPLANATION HERE As you can see in the three tables and three graphs above, the distribution of age and gender has been roughly equal.

6 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.