3B. Retrieval Practice and Learning

January 29, 2019

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 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 here: https://raw.githubusercontent.com/dlsun/data-science-book/master/data/KarpickeBlunt2011Replication/data.csv. - The codebook (explaining what the variables mean) is here: https://raw.githubusercontent.com/dlsun/data-science-book/master/data/KarpickeBlunt2011Replication/codebook.csv.

```
In [1]: %matplotlib inline
        import matplotlib.pyplot as plt
        import pandas as pd
        from altair import *

        df = pd.read_csv("https://raw.githubusercontent.com/dlsun/data-science-book/master/data
        df.head()
```

```
Out[1]:     ID  Age  Gender   Date.P1   Date.P2 Condition  IC.1  IC.2  Comp.1  Comp.2  \
        0  KB1   18  Female  11/21/16  11/28/16   Concept     1     1       1       1
        1  KB2   18    Male  11/21/16  11/28/16   Concept     1     1       1       1
        2  KB3   18    Male  11/21/16  11/28/16   Concept     1     1       1       1
        3  KB4   19  Female  11/21/16  11/28/16   Concept     1     1       1       1
        4  KB5   19  Female  11/22/16  11/29/16   Concept     1     1       1       1
```

```
      ...      Scorer.2.2  R2CS.avg  TS.1 Scorer.1.3  TS.2 Scorer.2.3  TS.avg  \
0     ...             NaN       NaN  0.36         NK  0.42         MS    0.39
1     ...             NaN       NaN  0.48         MS  0.36         NK    0.42
2     ...             NaN       NaN  0.08         NK  0.08         MS    0.08
3     ...             NaN       NaN  0.44         MS  0.42         NK    0.43
4     ...             NaN       NaN  0.26         NK  0.28         MS    0.27

   Exc.1 Exc.2  Collection
0      0     0           1
1      0     0           1
2      0     0           1
3      0     0           1
4      0     0           1

[5 rows x 35 columns]
```

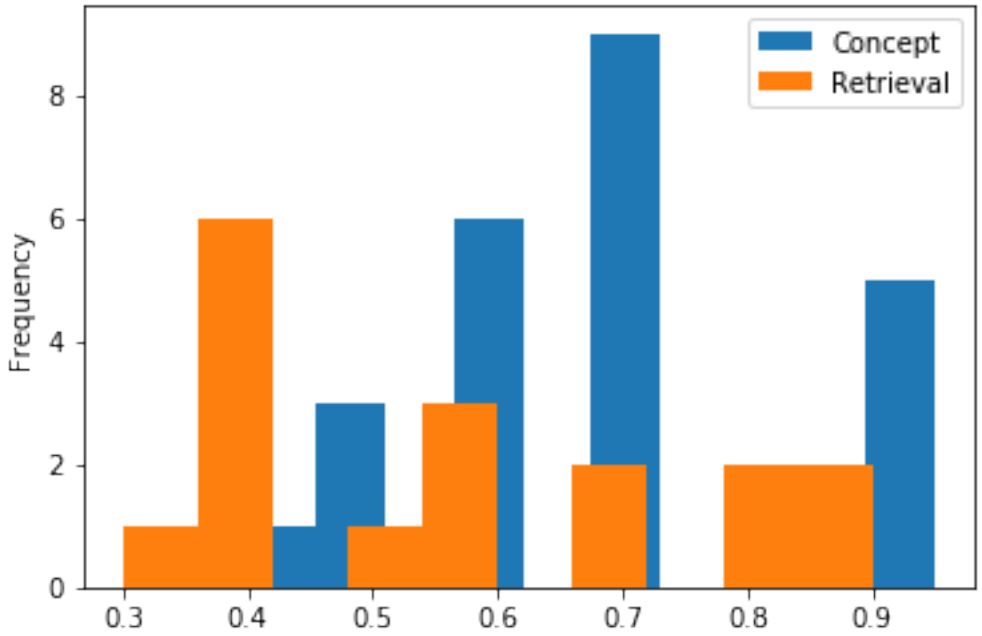
2 Question 1

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.

```
In [2]: df.groupby(["Condition"])["PR.2"].plot.hist(legend=True)
        df.groupby(["Condition"])["PR.2"].describe()
```

```
Out[2]:            count      mean       std  min  25%  50%    75%   max
        Condition
        Concept     26.0  0.690385  0.142842  0.4  0.6  0.7  0.775  0.95
        Retrieval   17.0  0.576471  0.195350  0.3  0.4  0.6  0.700  0.90
```



The subjects who made the concept maps felt like they learned more in the experiment. From the distribution of predicted scores, we see that the mean predicted score as a percentage is 12% higher in the concept map group than in the retrieval group. This 12% is greater than one whole letter grade. From the summary statistics, subjects in the concept maps group had a less variance in their predicted scores, and at each quartile, their predicted score is higher than that of the retrieval group. All of this indicates that perhaps the subjects in the concept maps group will do better than the subjects in the retrieval group.

In order to create separate histograms and summary statistics of the 2 groups, while still keeping all of the information on one plot, I used the groupby function to group the subjects by their Condition, or treatment group.

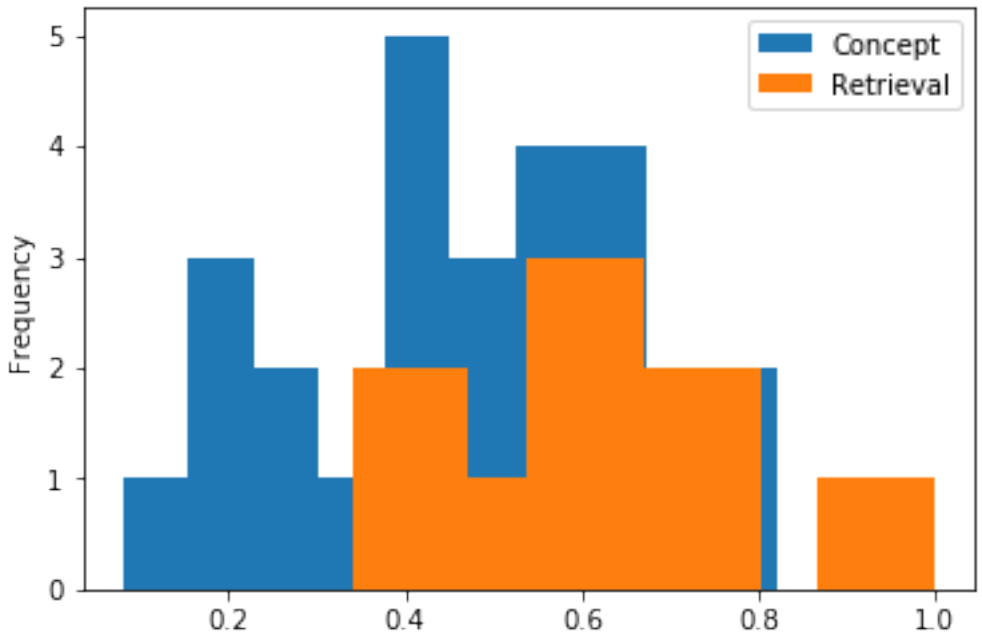
3 Question 2

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.

```
In [3]: df.groupby(["Condition"])["TS.1"].plot.hist(legend=True)
        df.groupby(["Condition"])["TS.1"].describe()
```

```
Out[3]:            count      mean       std   min    25%   50%   75%   max
        Condition
        Concept     26.0  0.470000  0.187467  0.08  0.365  0.48  0.59  0.82
        Retrieval   17.0  0.624706  0.178680  0.34  0.480  0.62  0.72  1.00
```



The subjects from the retrieval group on average scored 15% higher than subjects in the retrieval group. Here we see the opposite results from the prediction scores. According to the summary statistics, at every quartile the subjects in the retrieval group had a higher score as a percentage than the subjects in the concept map group. There was even a subject with a perfect score in the retrieval group whereas the highest score from the concept map group was only 82%.

I obtained this answer similarly to how I obtained the answer for question 1. I chose to use TS.1 as the test score because these were the test results from the original study whereas TS.2 was the results of a 2nd experiment in which the test was different from the first time.

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.)

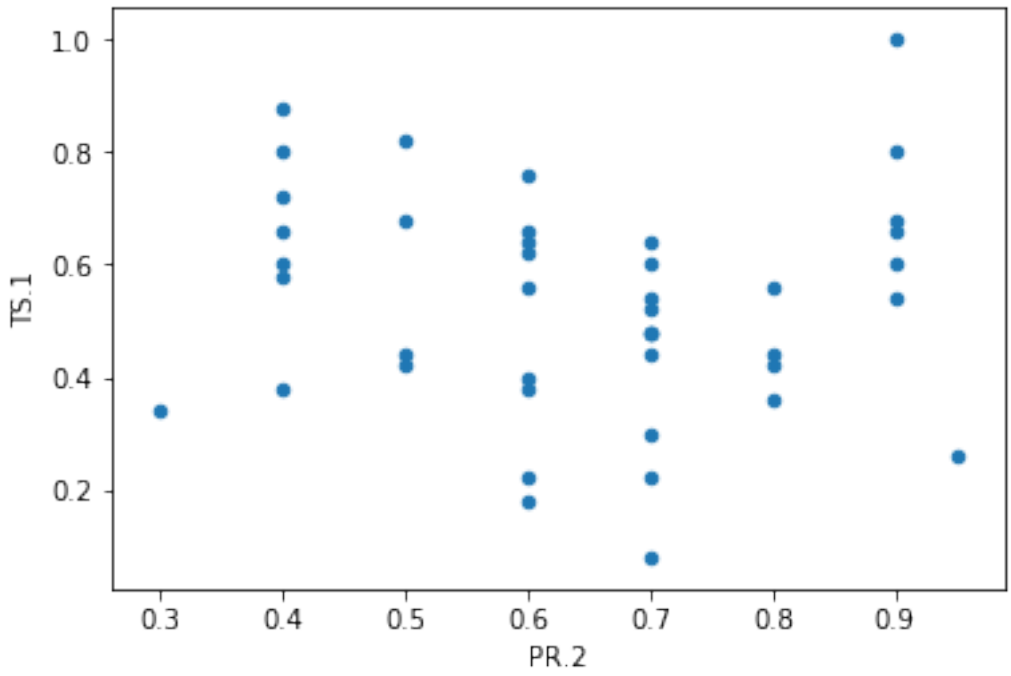
```
In [4]: df.plot.scatter("PR.2", "TS.1")

        [df["PR.2"].corr(df["TS.1"]),

        (df[df.Condition == "Concept"])["PR.2"].corr(
                (df[df.Condition == "Concept"])["TS.1"]),

        (df[df.Condition =="Retrieval"])["PR.2"].corr(
            (df[df.Condition == "Retrieval"])["TS.1"])]

Out[4]: [-0.061983386491153973, -0.064978193491088618, 0.24330632316924133]
```



From the scatter plot, it is difficult to say how well students were at predicting their score on the follow-up learning test. The combined correlation between the test results and predicted scores is practically zero. When we observe the correlation between test results and predicted scores by groups, we see a slightly different answer. For subjects in the concept map group, once again the correlation is nearly zero, meaning we cannot detect any significant pattern(s) in subjects' test scores vs predicted scores. However in the retrieval group, we see that test scores vs the predicted scores yields a correlation of 0.243 meaning there is slightly positive linear association in test scores and predicted scores.

I first made the scatterplot in hope there would be a strong linear association in the data, but when the correlation turned out to be practically zero, I thought it might be better to divide the data by the condition. From here, I used the boolean masks to filter the data down to one condition and obtain one correlation coefficient for that group.

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.

```
In [5]: df["Prediction"] = df["PR.2"]
        df["TestScore"] = df["TS.1"]

In [6]: df.Age = df.Age.astype(str)

        df.Age.value_counts()
```

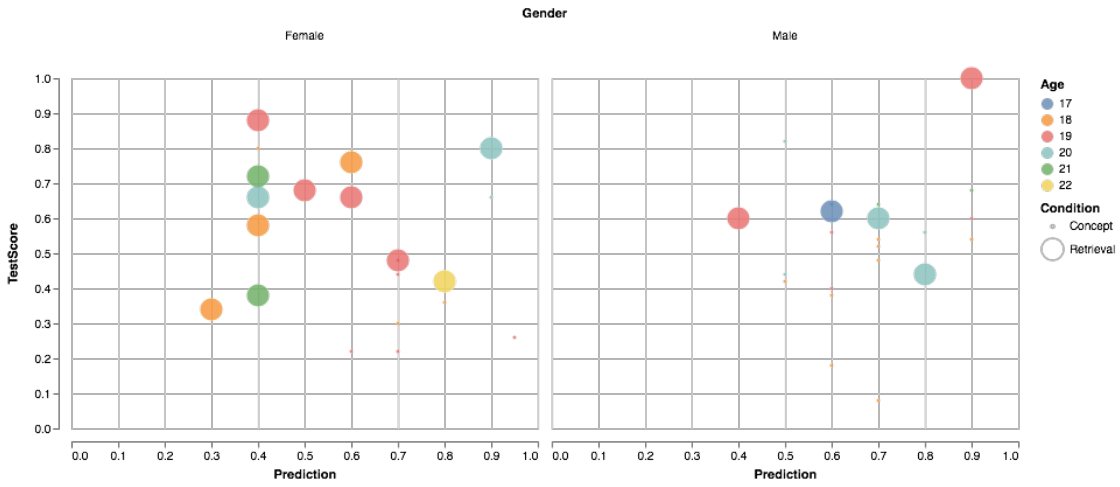
```
Out[6]: 18    14
        19    14
        20     9
        21     4
        17     1
        22     1
        Name: Age, dtype: int64
```

```
In [7]: df.Gender.value_counts()
```

```
Out[7]: Male      22
        Female    21
        Name: Gender, dtype: int64
```

```
In [8]: Chart(df).mark_circle().encode(
                x="Prediction",
                y="TestScore",
                color="Age",
                size="Condition",
                column="Gender"
        )
```

Out[8]:

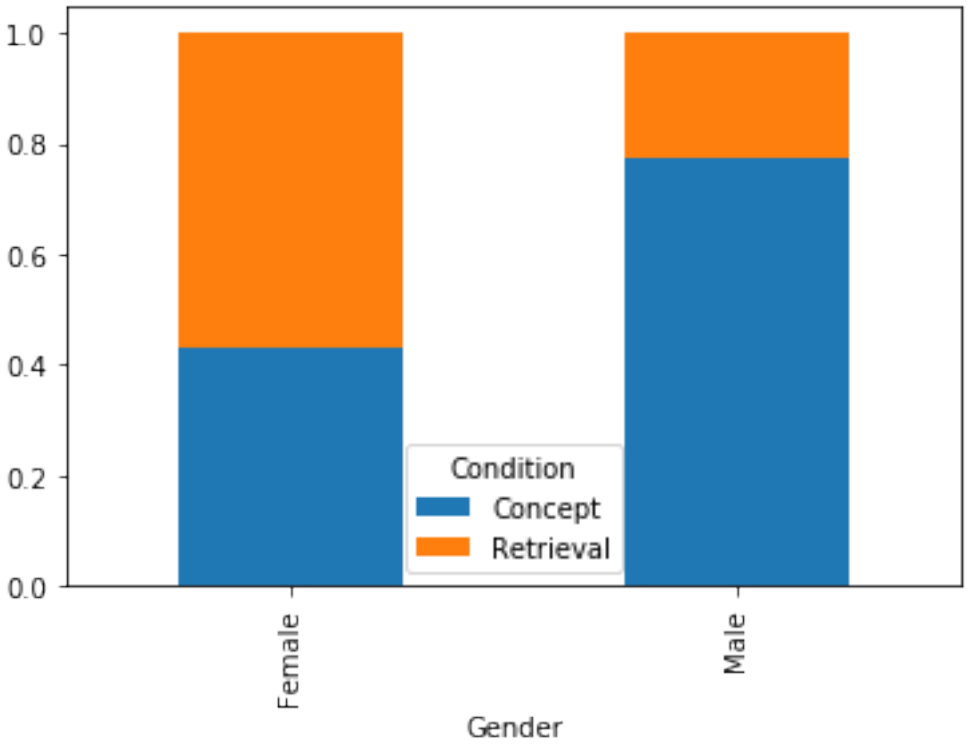



```
In [9]: counts = pd.crosstab(df.Gender, df.Condition)
        counts
```

```
Out[9]: Condition  Concept  Retrieval
        Gender
        Female           9         12
        Male            17          5
```

```
In [10]: condition_given_gender = counts.divide(counts.sum(axis=1), axis=0)
         condition_given_gender.plot.bar(stacked=True)
         condition_given_gender

Out[10]: Condition   Concept  Retrieval
         Gender
         Female     0.428571   0.571429
         Male       0.772727   0.227273
```



From this output above, we see that the number of males and females is equal, however, the distribution of Ages is not. We have many 18 and 19 year old participants, but only 1 22 year old student and 1 17 year old student. In the Altair plot, the scores appear to be roughly the same, but the males scores are more variable while the females predictions were lower. Lastly, if we look at the barplot, we see that the distribution of students to treatment groups given the student's gender are not equal. This violates the rule of independence that the 2 variables are independent if the conditional distribution of Y given X are all identical.

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 PolyLearn.