Data Literacy EDA

Adam Patterson

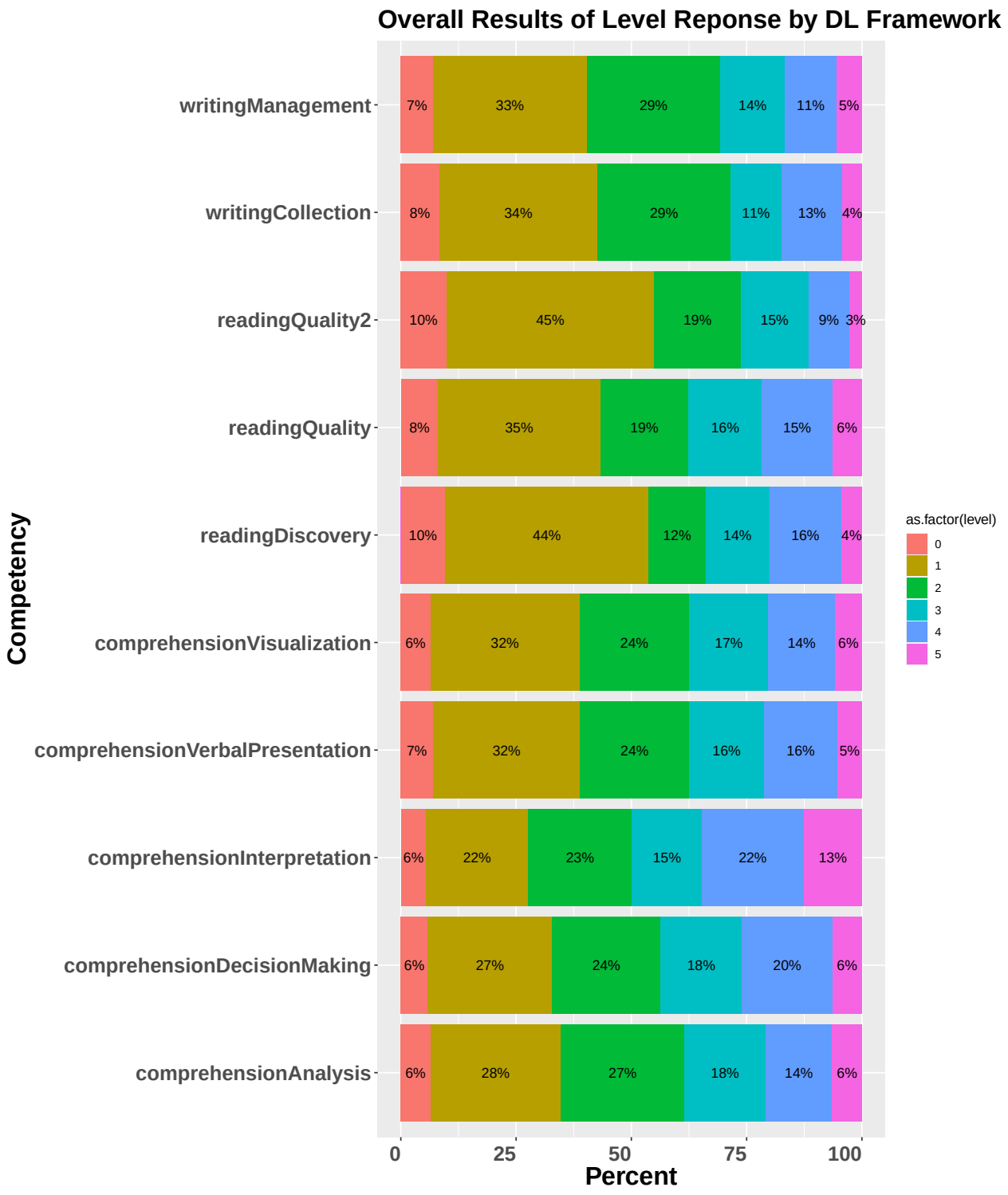
2023-03-02

Brief Visual EDA

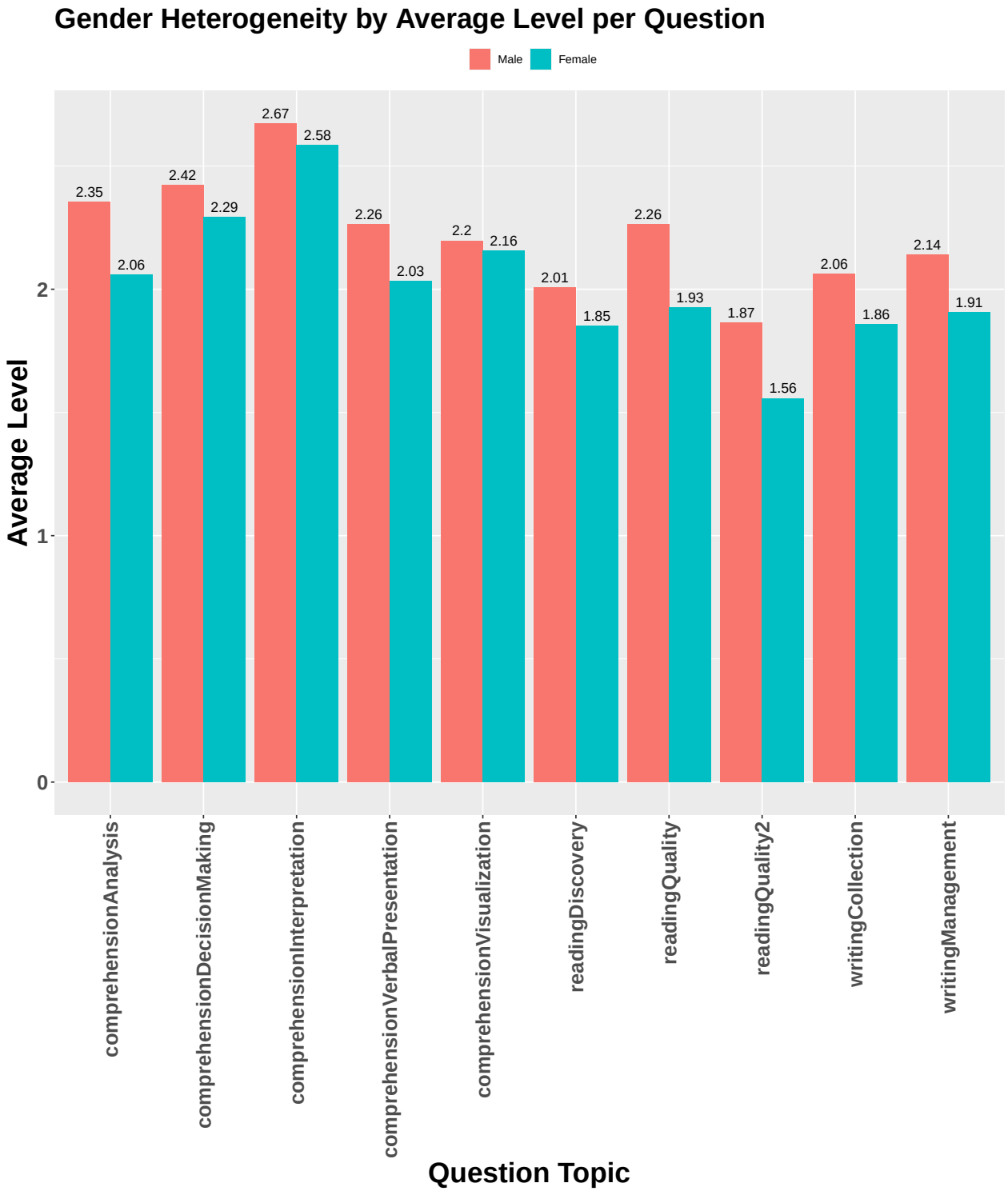
P value visualization uses bootstrapped simulated p values. The approx 0 p value across programmingHS and apCS are good to see. At least the results appear to be working intutitvely. The details are in my code script (not provided), but gender is the only variable that meets the minimum requirement of cell values for chi square. I tried these variables with fisher test but it asks to simulate a p value due to the same low frequency. I ran gender through fisher test (for smaller sample sizes and non large approximations) without simulating a p value, and then ran a simulated p value result. The result was a little different. I then bootstrapped the simulated value iterating 1000 times and the average was exact to the non-simulated version. I understand this does not mean anything significant because the data points were already there, and did not need to be simulated. The first result was a little different though. Nonetheless, I will look into the issue.

No data dictinary included in this analysis. Y values on page 5 graph are self explanatory with naming. The DL Framework competencies are named by branchCompetency . You will notice we had two readingQuality questions.

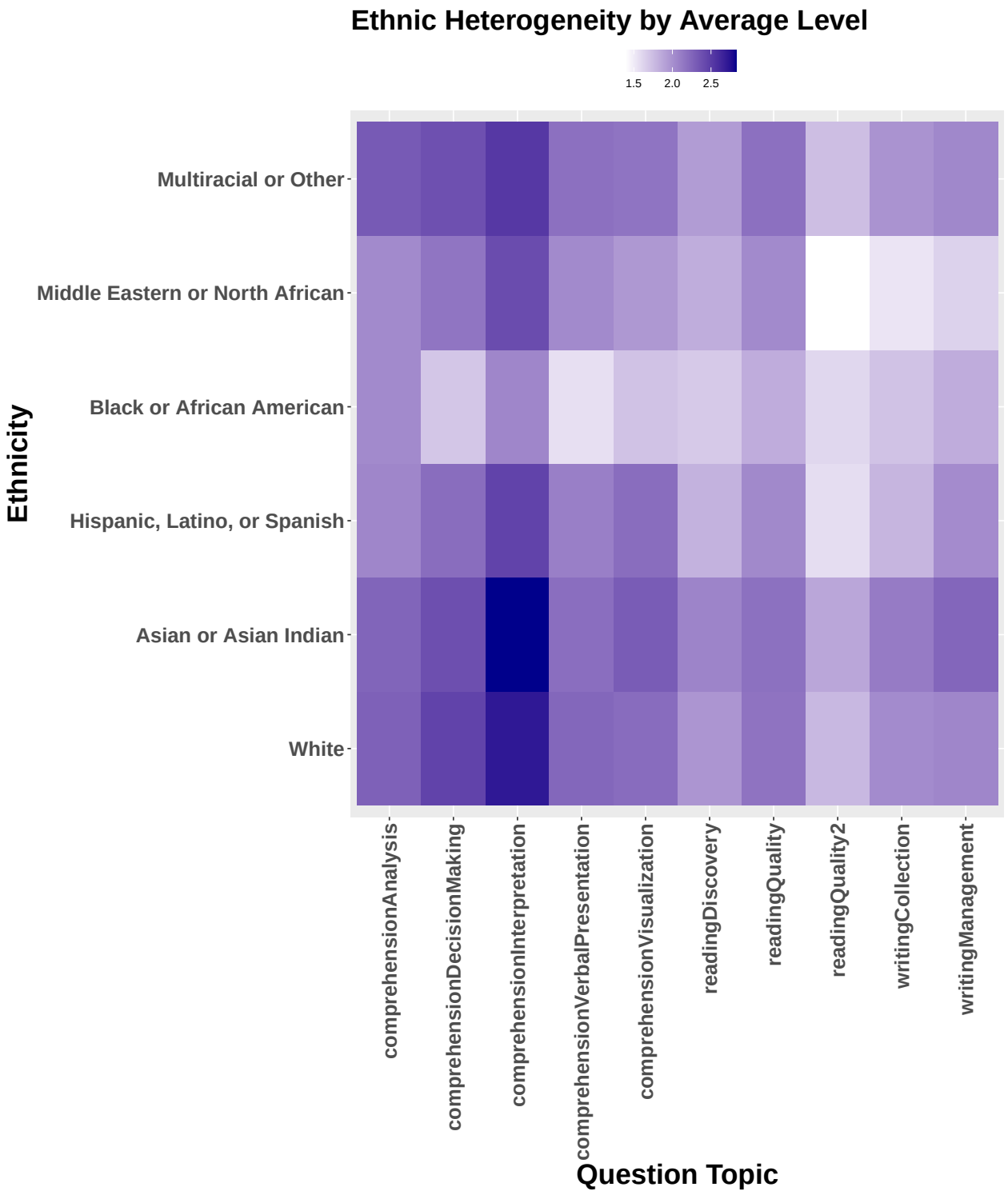
```
overall
```


Overall Results of Level Reponse by DL Framework
writingManagement
writingCollection
readingQuality2
readingQuality
readingDiscovery
comprehensionVisualization
comprehensionVerbalPresentation
comprehensionInterpretation
comprehensionDecisionMaking
comprehensionAnalysis
7% 33% 29% 14% 11% 5%
8% 34% 29% 11% 13% 4%
10% 45% 19% 15% 9% 3%
8% 35% 19% 16% 15% 6%
10% 44% 12% 14% 16% 4%
6% 32% 24% 17% 14% 6%
7% 32% 24% 16% 16% 5%
6% 22% 23% 15% 22% 13%
6% 27% 24% 18% 20% 6%
6% 28% 27% 18% 14% 6%
Competency
0 25 50 75 100
Percent
as.factor(level)
0
1
2
3
4
5

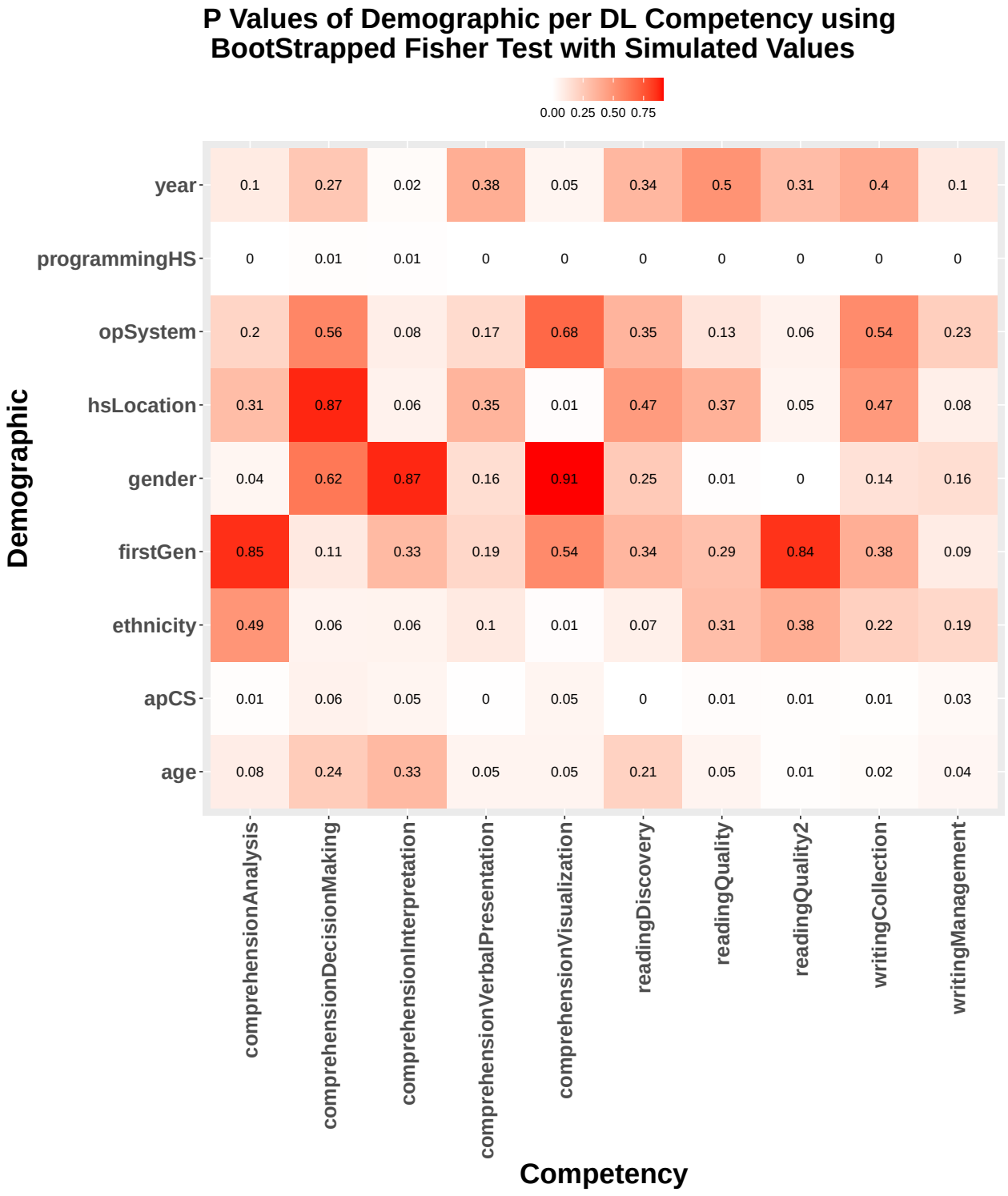
```
gender
```


Gender Heterogeneity by Average Level per Question
Male
Female
Average Level
0
1
2
2.35
2.06
2.42
2.29
2.67
2.58
2.26
2.03
2.2
2.16
2.01
1.85
2.26
1.93
1.87
1.56
2.06
1.86
2.14
1.91
comprehensionAnalysis
comprehensionDecisionMaking
comprehensionInterpretation
comprehensionVerbalPresentation
comprehensionVisualization
readingDiscovery
readingQuality
readingQuality2
writingCollection
writingManagement
Question Topic

```
ethnicity
```


Ethnic Heterogeneity by Average Level
1.5
2.0
2.5
Ethnicity
Multiracial or Other
Middle Eastern or North African
Black or African American
Hispanic, Latino, or Spanish
Asian or Asian Indian
White
comprehensionAnalysis
comprehensionDecisionMaking
comprehensionInterpretation
comprehensionVerbalPresentation
comprehensionVisualization
readingDiscovery
readingQuality
readingQuality2
writingCollection
writingManagement
Question Topic

```
pVal
```


P Values of Demographic per DL Competency using BootStrapped Fisher Test with Simulated Values
0.00 0.25 0.50 0.75
Demographic
year 0.1 0.27 0.02 0.38 0.05 0.34 0.5 0.31 0.4 0.1
programmingHS 0 0.01 0.01 0 0 0 0 0 0 0
opSystem 0.2 0.56 0.08 0.17 0.68 0.35 0.13 0.06 0.54 0.23
hsLocation 0.31 0.87 0.06 0.35 0.01 0.47 0.37 0.05 0.47 0.08
gender 0.04 0.62 0.87 0.16 0.91 0.25 0.01 0 0.14 0.16
firstGen 0.85 0.11 0.33 0.19 0.54 0.34 0.29 0.84 0.38 0.09
ethnicity 0.49 0.06 0.06 0.1 0.01 0.07 0.31 0.38 0.22 0.19
apCS 0.01 0.06 0.05 0 0.05 0 0.01 0.01 0.01 0.03
age 0.08 0.24 0.33 0.05 0.05 0.21 0.05 0.01 0.02 0.04
comprehensionAnalysis
comprehensionDecisionMaking
comprehensionInterpretation
comprehensionVerbalPresentation
comprehensionVisualization
readingDiscovery
readingQuality
readingQuality2
writingCollection
writingManagement
Competency