Spatial Reasoning and Statistical Graphics

Susan VanderPlas, Heike Hofmann

October 11, 2014

1 Introduction

Relevant literature:

- Shah and Carpenter (1995) showed that spatial ability was not correlated with accuracy on a simple two-dimensional line graph description task, but that mathematical ability was correlated with accuracy.

- Just and Carpenter (1985) showed that high-spatial-ability viewers used different rotation strategies than low-spatial-ability viewers when asked to whether three-dimensional alphabet cubes were the same.

- Hofmann et al. (2012) for lineup stimuli and general lineup performance

Lineups depend on the ability to search for a signal amid distractors (Visual Search Task) and the ability to infer patterns from stimuli (Pattern Recognition task). Some lineups (polar coords) also depend on the ability to mentally rotate stimuli (spatial rotation task) and mentally manipulate graphs (paper folding task). By breaking the lineup task down into component parts, we can correlate lineup performance with similar cognitive factor tests to determine where additional variation in skill level factors into performance differences. In addition, we can correlate previous experiences (science-based major, research experience, Auto-CAD skills) with performance to explore the effect that participant experience has on lineup performance.

2 Methods

Participants will complete the following tasks (sample pictures included, full stimuli set will be added to the appendix once testing is complete). Tasks are designed so that participants are under time pressure; they are not expected to complete all of the problems in each section. This provides more discrimination between high scorers and prevents score compression at the top of the range.

- Visual Search Task: designed to test participants' ability to find a target stimulus in a field of distractors. An example is shown in figure 1.

- Paper Folding Task: tests participants' ability to visualize and mentally manipulate figures in three dimensions. Associated with the ability to extrapolate symmetry and reflection over multiple steps. An example is shown in figure 2.

- Card Rotation Task: tests participant's ability to rotate objects in two dimensions to distinguish between left-hand and right-hand versions of the same figure. Tests spatial reasoning ability and mental rotation skills. An example is shown in figure 3.

- Figure Classification Task: tests participant's ability to extrapolate rules from provided figures. This task is associated with visual reasoning capabilities and we expect that it should correlate with the ability to pick out a signal plot from a lineup. An example is shown in figure 4.

Figure 1: Visual Search Task. Participants are instructed to find the plot numbered 1-24 which matches the plot labeled "Target". Participants will complete up to 25 of these tasks in 5 minutes.

Figure 2: Paper Folding Task. Participants are instructed to pick the figure matching the sequence of steps shown in the left-hand figure. Participants will complete up to 20 of these tasks in 6 minutes.

Figure 3: Card Rotation Task. Participants mark each figure on the right hand side as either the same or different than the figure on the left hand side of the dividing line. Participants will complete up to 20 of these tasks (each consisting of 8 figures) in 6 minutes.

Figure 4: Figure Classification Task. Participants classify each figure in the second row as belonging to group 1, 2, or 3 (if applicable). Participants will complete up to 14 of these tasks (each consisting of 8 figures to classify) in 8 minutes.

Between cognitive tasks, participants will also complete three blocks of 20 lineups each. These lineups have been previously tested (Hofmann et al., 2012) and include some null lineups (i.e. lineups without a target plot). Participants have 5 minutes to complete each block of 20 lineups. Figure 5 shows a sample lineup of box plots.

In addition to these tests, participants will complete a questionnaire which includes questions about colorblindness, mathematical background, self-perceived verbal/mathematical/artistic skills, time spent playing video games, and undergraduate major. These questions are designed to assess different factors which may influence a participant's skill at reading graphs and performing spatial tasks.
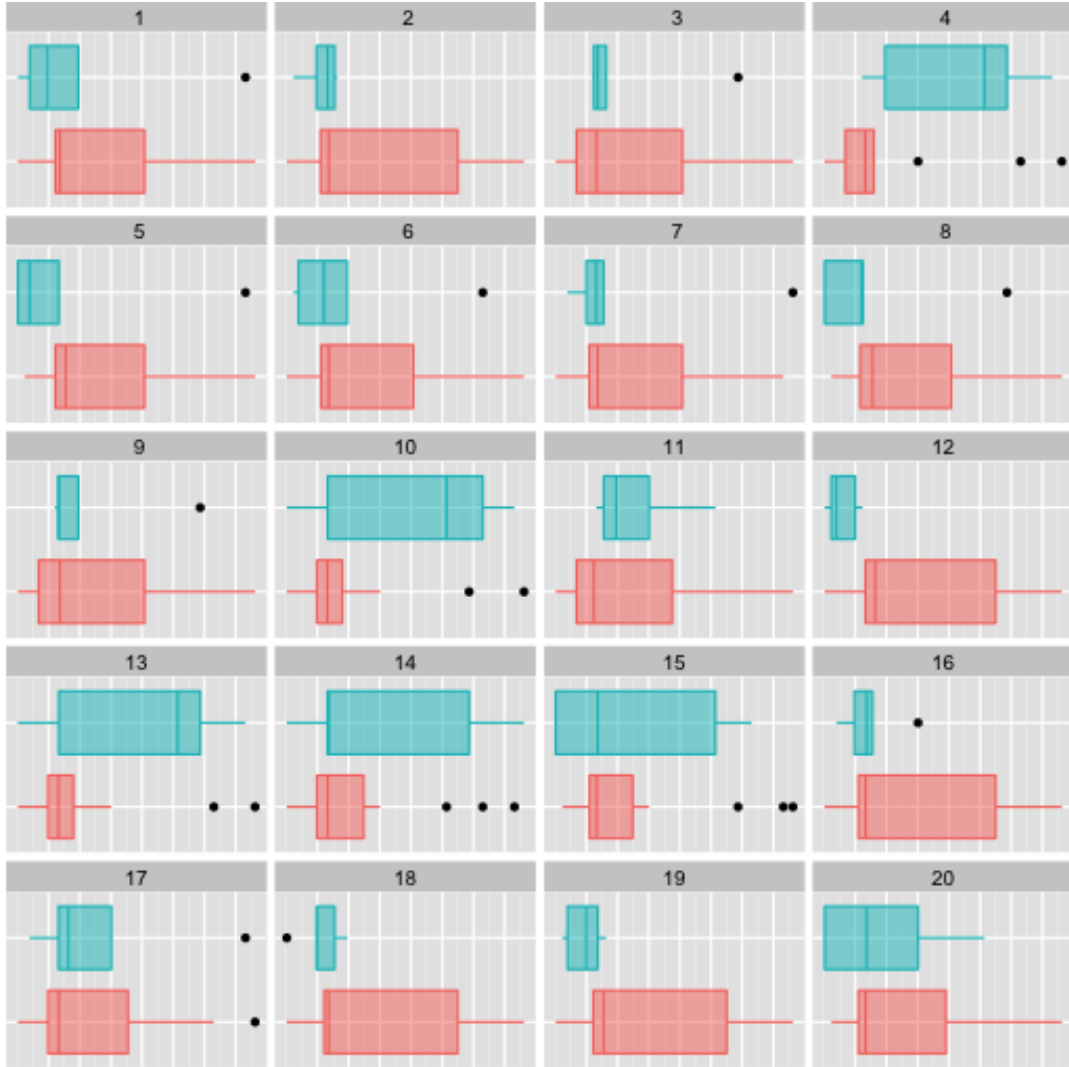
Figure 5: A sample lineup. Participants are instructed to choose the plot which appears most different from the others. In this lineup, plot 13 is the target.

3 Results

No woulds, shoulds or coulds

Scoring for all of the tests was calculated such that with random guessing, the expected score on a problem would be 0; thus, for a test with possible answers A, B, C, D, and E, the score would be calculated by subtracting $1/4 * \#$ wrong answers from the total number of correct answers.

re-phrasing the above sentence:

Scoring of all test results was done such that random guessing leads to an expected value of 0; thus for a test consisting of multiple choice questions with k suggested answers with a single correct answer each, the score is calculated as

$$\#\text{total correct answers} - 1/k \cdot \#\text{wrong answers}. \tag{1}$$

This allows us to compare each participant's score in light of how many problems were attempted as well as the number of correct responses.

Combining accuracy and speed into a single number does not only make a comparison of test scores easier,

this scoring mechanism is also used on many standardized tests, such as the SAT, ACT, and the battery of psychological tests which

parts

of this test are drawn from.

We need a few citations here.

The next step is an overview of the possible ranges of the test scores in our study, because that is driven by the number of tests shown. It would be good to scale everything, so that we have the same theoretical range. That will actually allow a comparison across different tests.

```
library(plyr)
ldply(ans.summary[,c("lineup","card_rot", "fig_class", "folding", "vis_search")], c(mean=mean,sd= sd))

##           .id   mean     sd
## 1      lineup  5.170  2.117
## 2    card_rot 17.806 21.596
## 3   fig_class 69.583 21.243
## 4     folding  5.729  2.346
## 5  vis_search 22.692  2.578
```

If test scores have the same ranges, we could include a mini table of means and standard deviations of the four tests, and discuss whether these findings are consistent to how people usually score (except for the females we have about the same population).

Split the next sentence up into two paragraph: what is shown in each of the figures, what are the immediate conclusions? chuck out the gaming hours from figure 6, that's a demograph, so move it to the other plot (I am aware that nine demographics are easier to display than ten, but content trumps). What happened to the arts skills? I seem to have a positive association now - why is my data different from yours?

Results are presented graphically in figures 6 and 7.

```
cor(ans.summary[,c("lineup","card_rot", "fig_class", "folding", "vis_search")])

##           lineup card_rot fig_class folding vis_search
## lineup    1.0000   0.5154    0.6970  0.5116     0.4130
## card_rot  0.5154   1.0000    0.5392  0.7414     0.6525
## fig_class 0.6970   0.5392    1.0000  0.5775     0.4172
## folding   0.5116   0.7414    0.5775  1.0000     0.4280
## vis_search 0.4130  0.6525    0.4172  0.4280     1.0000

# using scaled version right now, should be changed to unscaled once the scores are internally scaled.
pca <- prcomp(ans.summary[,c("lineup","card_rot", "fig_class", "folding", "vis_search")], scale=T)
summary(pca)

## Importance of components:
##                          PC1   PC2   PC3    PC4    PC5
## Standard deviation     1.791 0.865 0.740 0.5472 0.4420
## Proportion of Variance 0.642 0.150 0.110 0.0599 0.0391
## Cumulative Proportion  0.642 0.791 0.901 0.9609 1.0000

screeplot(pca)
```
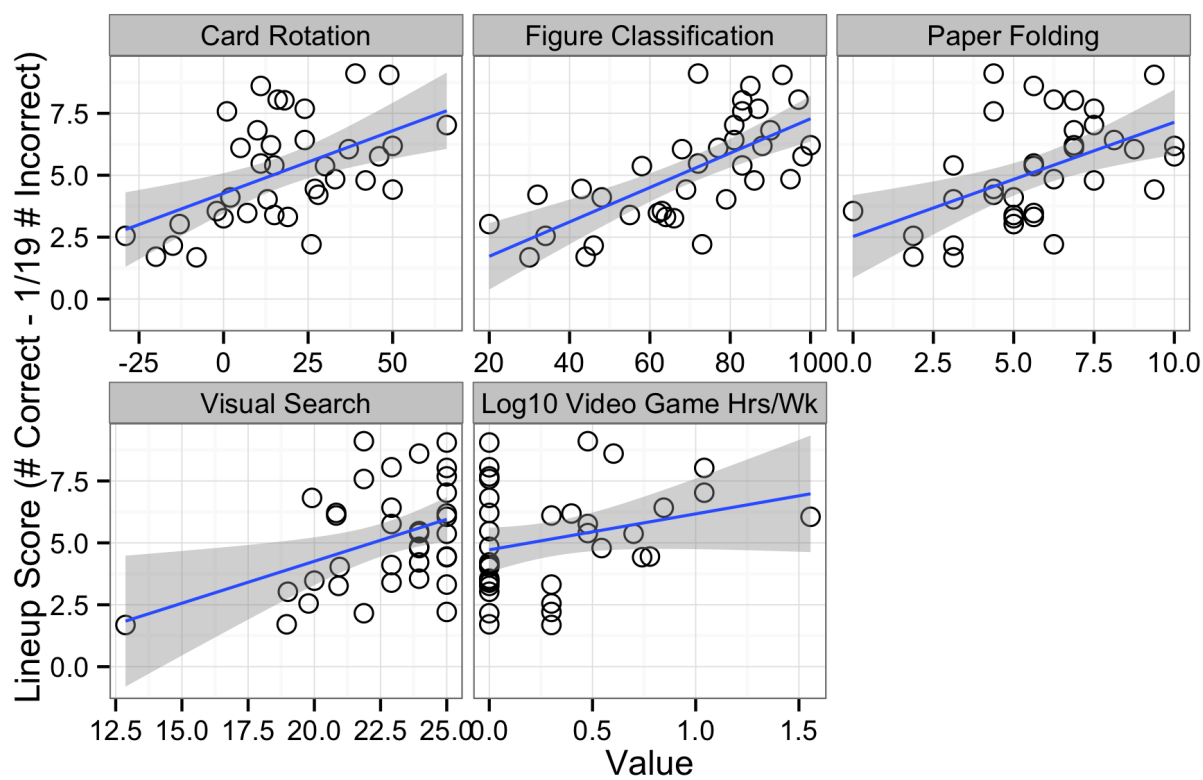

pca
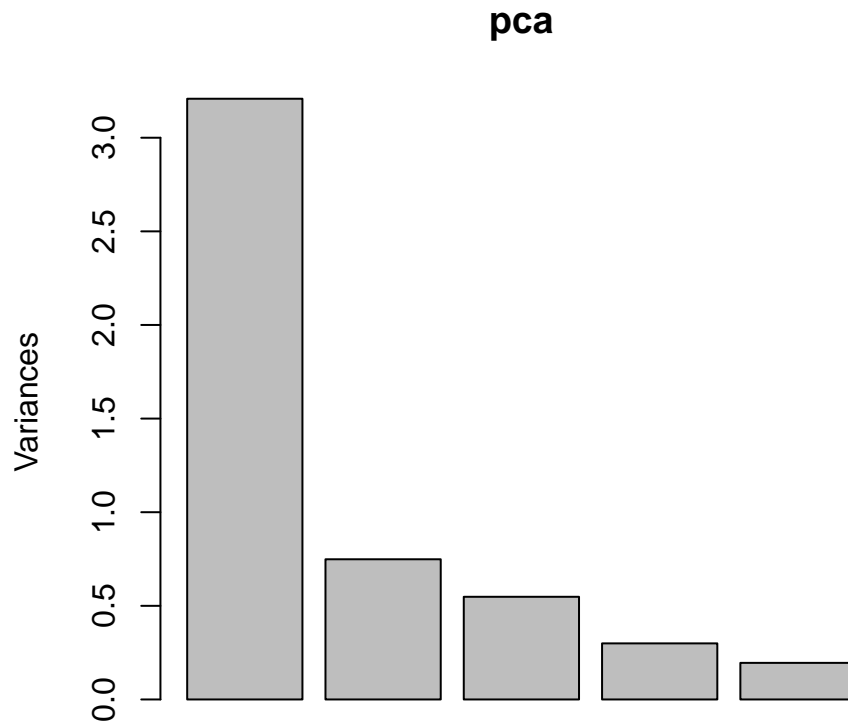
Figure 6: Preliminary results of continuous variables compared to lineup score.
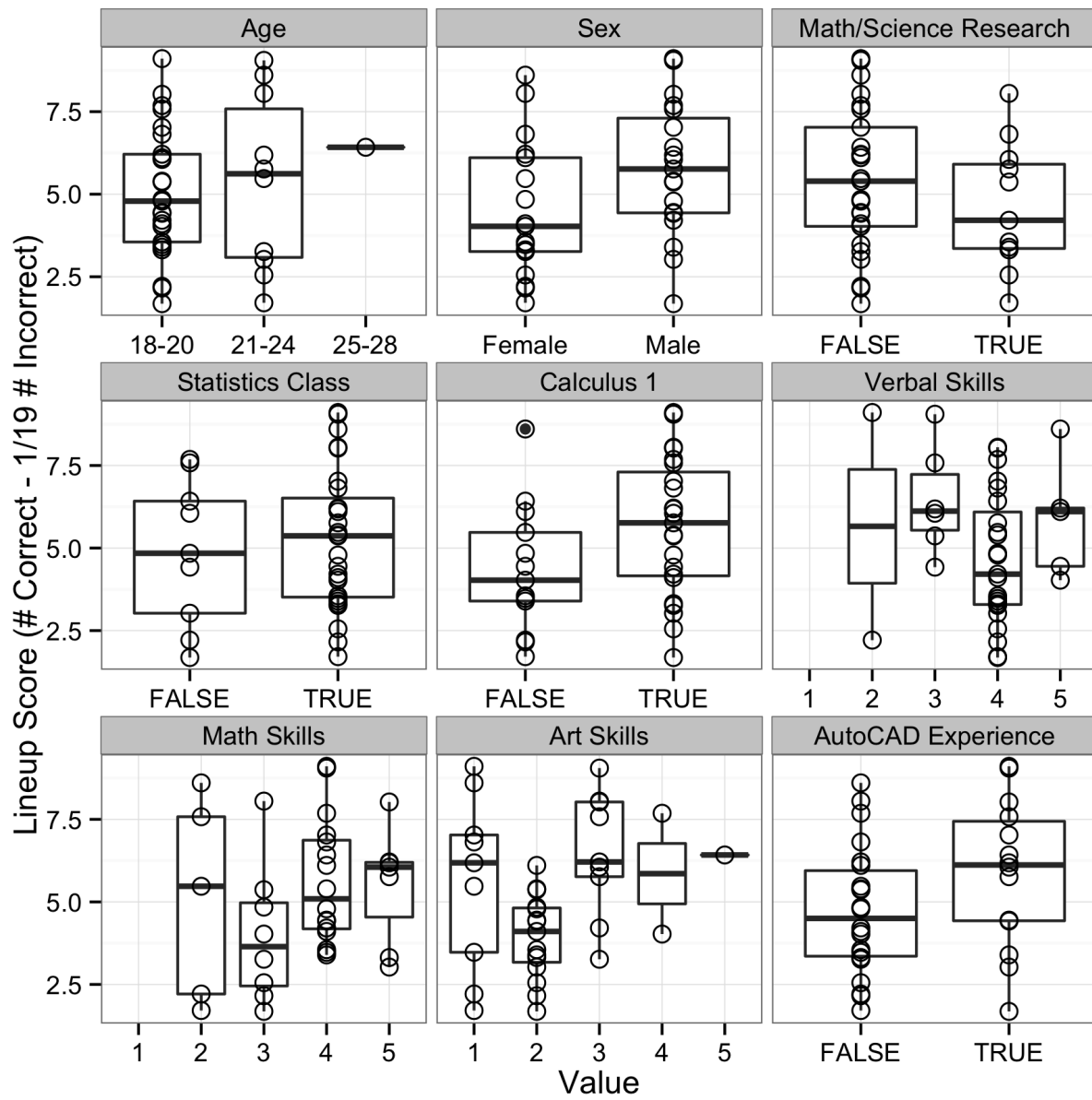
Figure 7: Preliminary results of categorical variables compared to lineup score.

All results and data shown here are done in accordance with IRB # XXX?

References

Hofmann, H., Follett, L., Majumder, M., and Cook, D. (2012). Graphical tests for power comparison of competing designs. Visualization and Computer Graphics, IEEE Transactions on, 18(12):2441–2448.

Just, M. A. and Carpenter, P. A. (1985). Cognitive coordinate systems: accounts of mental rotation and individual differences in spatial ability. Psychological review, 92(2):137.

Shah, P. and Carpenter, P. A. (1995). Conceptual limitations in comprehending line graphs. Journal of Experimental Psychology: General, 124(1):43.

Appendix

T-tests of results for Hillary and Stephanie:

```
t.test(ans.summary$card_rot[1:18], ans.summary$card_rot[-c(1:18)])

##
##  Welch Two Sample t-test
##
## data:  ans.summary$card_rot[1:18] and ans.summary$card_rot[-c(1:18)]
## t = 1.384, df = 33.37, p-value = 0.1756
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -4.619 24.285
## sample estimates:
## mean of x mean of y
##     22.72     12.89

t.test(ans.summary$folding[1:18], ans.summary$folding[-c(1:18)])

##
##  Welch Two Sample t-test
##
## data:  ans.summary$folding[1:18] and ans.summary$folding[-c(1:18)]
## t = 1.735, df = 33, p-value = 0.09202
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.2275  2.8664
## sample estimates:
## mean of x mean of y
##     6.389     5.069

t.test(ans.summary$lineup[1:18], ans.summary$lineup[-c(1:18)])

##
##  Welch Two Sample t-test
##
## data:  ans.summary$lineup[1:18] and ans.summary$lineup[-c(1:18)]
## t = 2.582, df = 33.25, p-value = 0.01441
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
```

```
##  0.3589 3.0213
## sample estimates:
## mean of x mean of y
##     6.015     4.325
```

```r
t.test(ans.summary$vis_search[1:18], ans.summary$vis_search[-c(1:18)])
```

```
##
##  Welch Two Sample t-test
##
## data:  ans.summary$vis_search[1:18] and ans.summary$vis_search[-c(1:18)]
## t = 0.9918, df = 29.39, p-value = 0.3294
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.9046  2.6099
## sample estimates:
## mean of x mean of y
##     23.12     22.27
```