The study of false memories

A replication and extension of
Roediger and McDermott

Roediger and McDermott

Meta-theoretical question

(1) memory as photograph versus memory as reconstruction

(2) “recovered” childhood memories of trauma versus “false” memories

(3) legal testimony of accuracy of memory

Roediger and McDermott-background

Prior work

(1) memory distortions over time -- Bartlett

(2) reconstructive memory -- Loftus

(3) low error rates in recognition memory -- Underwood

(4) intrusions in free recall -- Deese

Roediger and McDermott

Alternative explanations for memory effects

(1) connection strength models of memory

(2) network models of association

Theoretical statement

(1) not testing theory but rather testing phenomenon

(2) need to get a robust measure of false memory in order to study it

Roediger and McDermott Study 1

Materials

(a) 6 lists of 12 words with high associates of 6 target lures

(b) recognition list

- i) 12 studied words
- ii) 6 target lures
- iii) 12 weakly related
- iv) 12 unrelated

Procedure

(a) verbal presentation of each list

(b) free recall after each list

(c) recognition 2 minutes after all lists had been presented

Results

(a) recall shows serial position effects

(b) intrusion errors almost as strong as low point of serial position

(c) recognition errors are frequent

Roediger and McDermott Study 2

Materials

(a) 16 lists

procedure

results

205 Replication of R & M

- Can we get the same effect?
- What is the effect of presentation (study) time
- What is the effect of recall period?
- Do study and recall times have parallel effects on false memory effect?
- Do real memories and false memories behave the same way?

Method

- Materials
  - 16 lists of 15 words (R&M lists 1-16)
  - 16 recall/math lists
  - 3 pages of math problems
  - Recognition list (3 words from each of lists 1-24 plus the "prime" from all lists)

Method

- Procedure
  - Study words presented for 2 or 3 seconds
  - Recall interval of 45 or 90 seconds
  - Instructions to “recall” or to do math problems
  - 3 minute filler task
  - 3 minute recognition task

How to study several within subject variables at the same time

- Counterbalancing to avoid confounding
- Conditions crossed with conditions
- Conditions balanced across orders

Design for Class Data

List	Time	Recall	A/B
1	2	45	Recall
2	3	90	Math
3	3	45	Math
4	2	90	Recall
5	3	90	Recall
6	2	45	Math
7	2	90	Math
8	3	45	Recall
9	3	90	Math
10	2	45	Recall
11	2	90	Recall
12	3	45	Math
13	2	45	Math
14	3	90	Recall
15	3	45	Recall
16	2	90	Math

Results - Descriptive

- Descriptive statistics vs. Inferential stats
- Describe the results --
  - Say it in words

    Say it in pictures (figures)

    Say it in numbers
- Inferential: What is the likelihood that the results could happen by chance?

  Estimate a parameter and give confidence intervals for that parameter

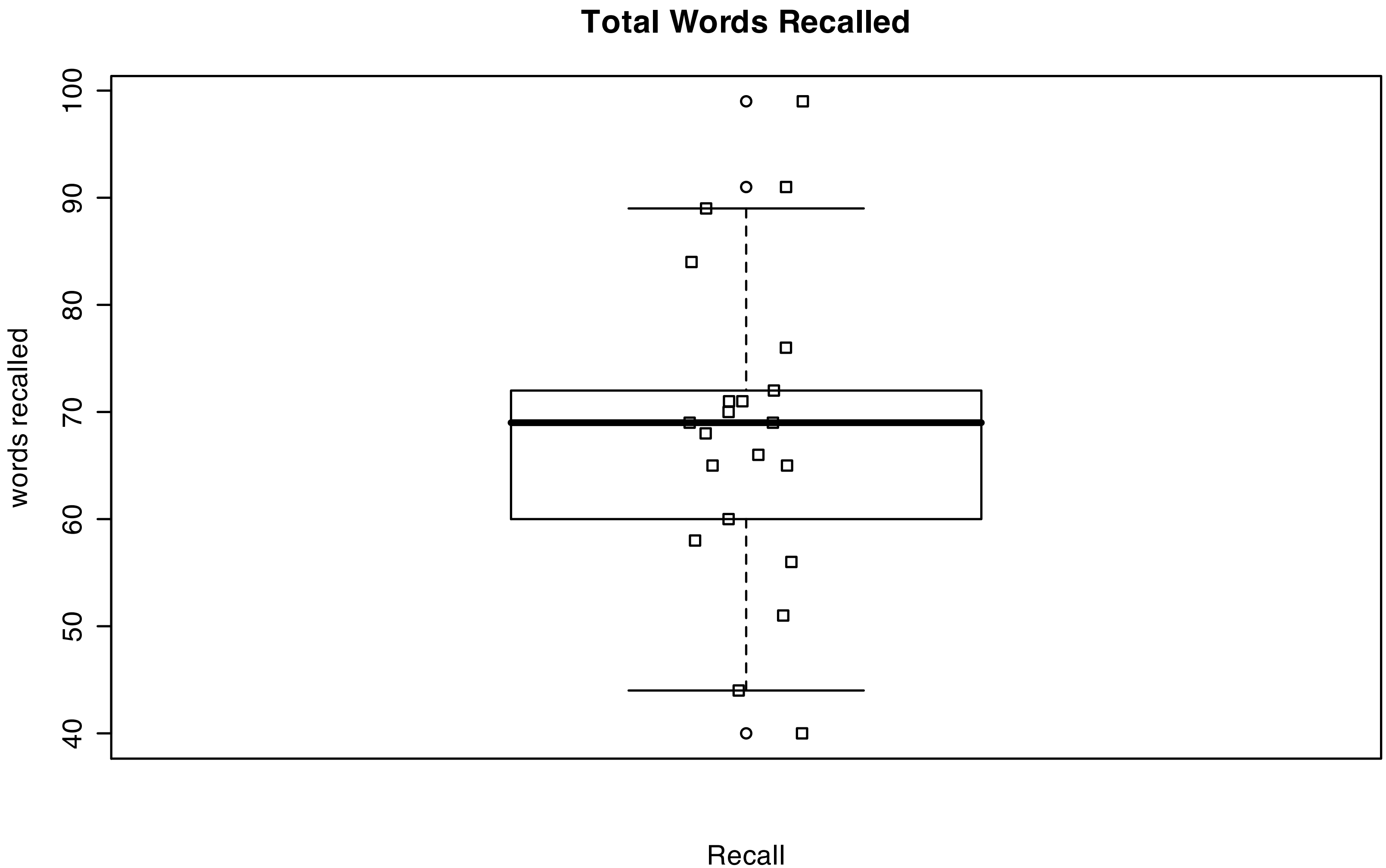
Results - selective summary

- No need to report every analysis, just the ones that tell the important story
- Think about how to aggregate the data to best summarize it
- Transforms of data to make more understandable
  - e.g., percent correct rather than raw number
- Story must be truth
  - don't hide "inconvenient data"
  - assume someone else will want to analyse your data

Data= Model + Error

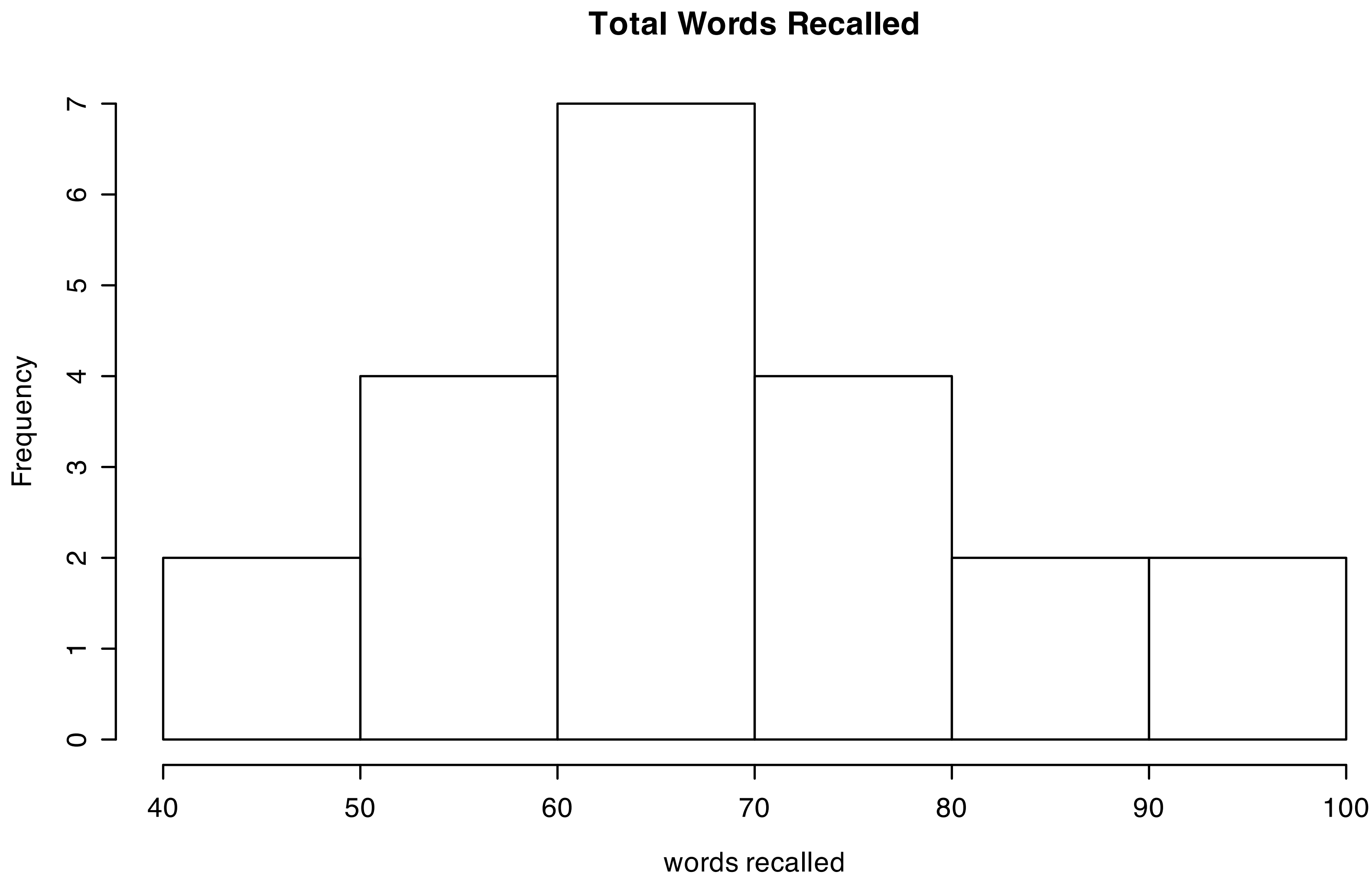
- The process of science is improve the model and reduce the error
- Models are progressively more complicated
- Consider the recall data:
  - Model 0: Data
  - Model 1: Data = Mean + Error
  - Model 2: Data = $Position_i$ + Error

Model 1:Median + Error



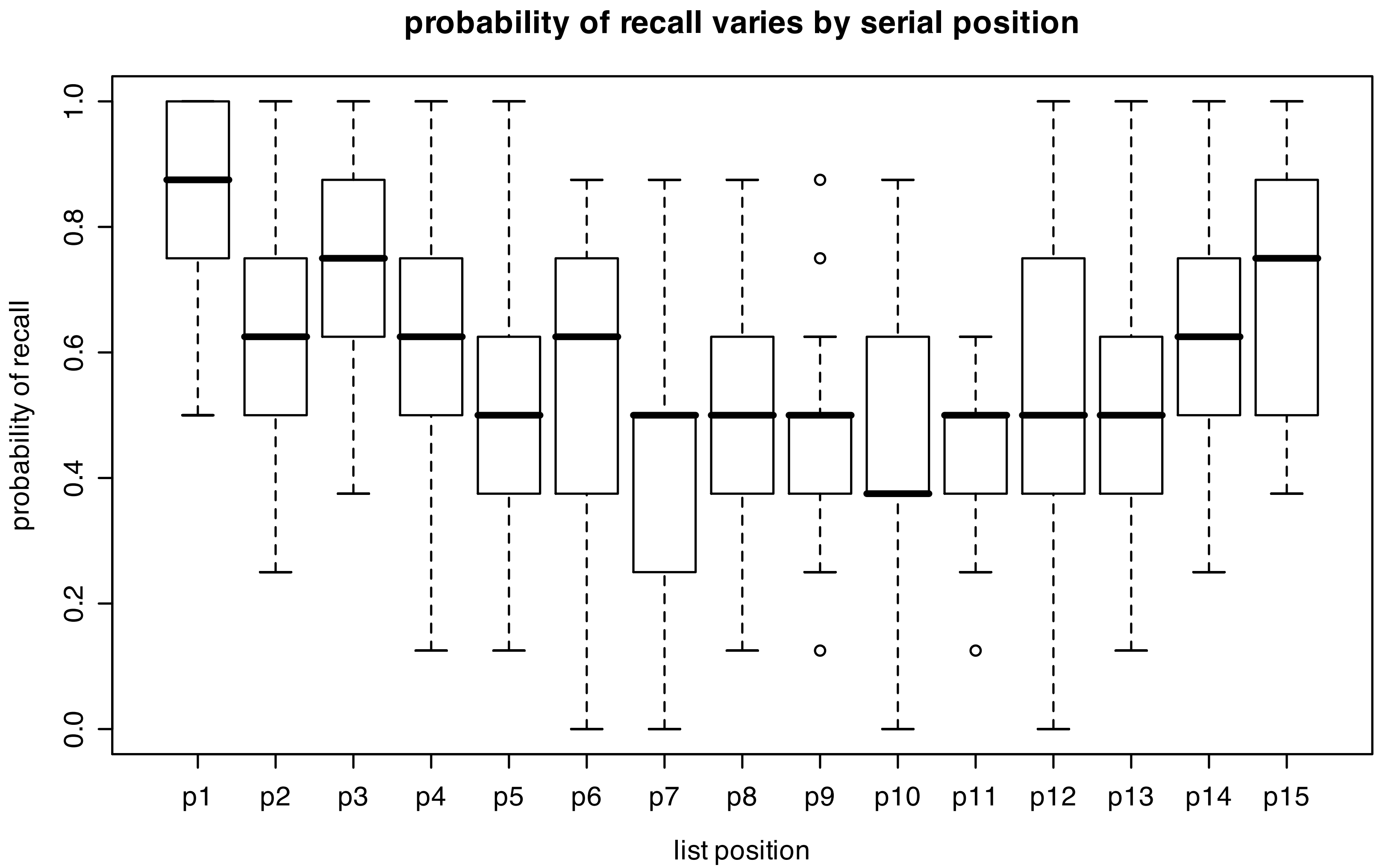
Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
40.00	60.00	69.00	68.29	72.00	99.00

Histograms show frequency of score



Model 2:
Recall varies by Serial Position + Error

```
boxplot(position/8,ylab="probability of recall",xlab="list position", main="probability of recall varies by serial position")
```

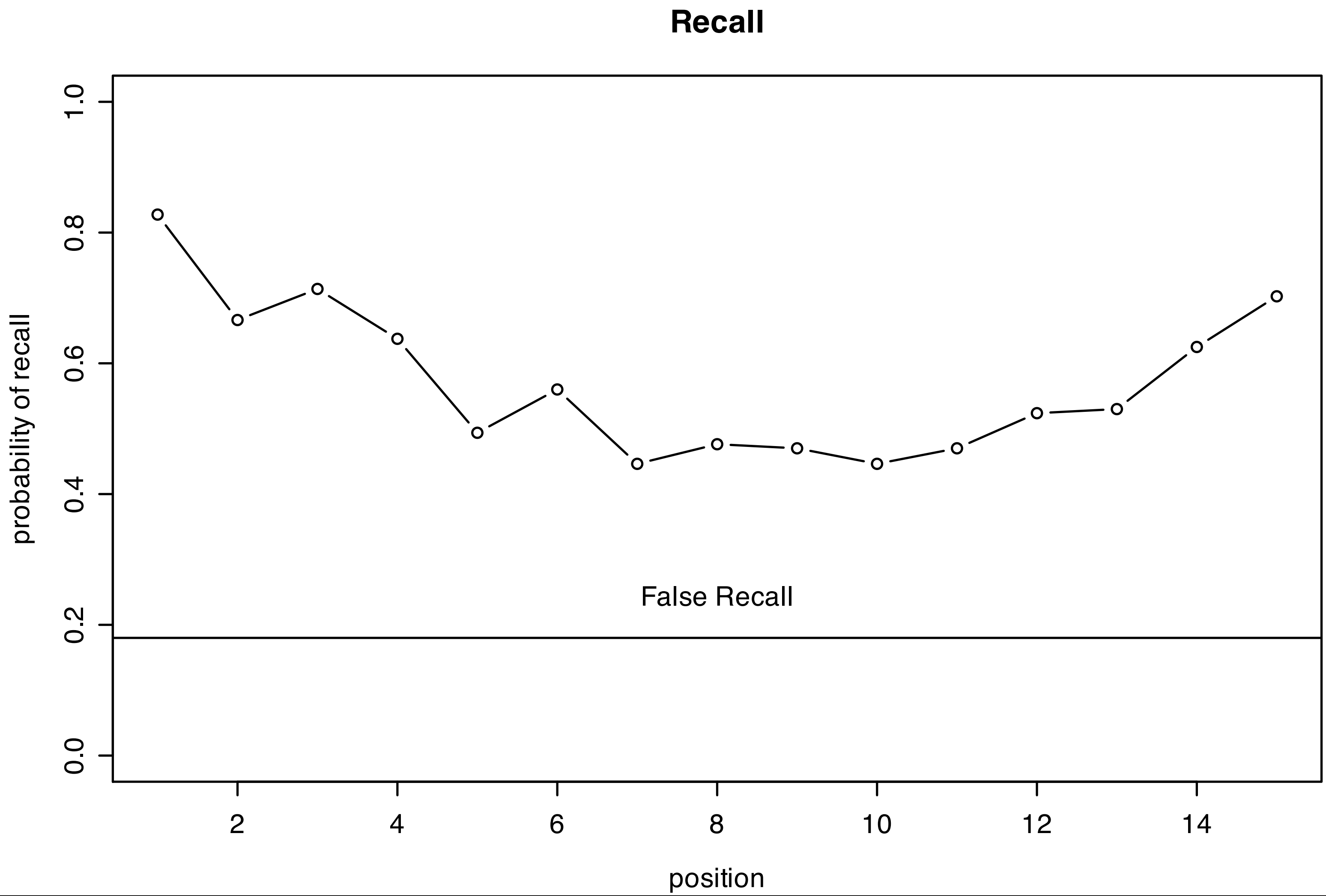


Graphics vs. tables

	mean	sd	skew	n	median	se
p1	6.62	1.28	-0.52	21	7	0.28
p2	5.33	1.59	-0.10	21	5	0.35
p3	5.71	1.31	-0.51	21	6	0.29
p4	5.10	1.79	-0.39	21	5	0.39
p5	3.95	1.80	0.60	21	4	0.39
p6	4.48	1.99	-0.67	21	5	0.43
p7	3.57	1.72	-0.03	21	4	0.38
p8	3.81	1.36	0.33	21	4	0.30
p9	3.76	1.41	0.20	21	4	0.31
p10	3.57	1.91	0.10	21	3	0.42
p11	3.76	1.00	-0.99	21	4	0.22
p12	4.19	1.89	-0.01	21	4	0.41
p13	4.24	1.73	0.15	21	4	0.38
p14	5.00	1.48	0.09	21	5	0.32
p15	5.62	1.63	-0.20	21	6	0.36

Recall by serial position

```
recall.stats <-   describe(position)
plot(recall.stats$mean/8,ylim=c(0,1),type="b",ylab="probability of recall", xlab="position", main="Recall")
abline(h=.18)
 text(8,.25,"False Recall")
```



Results

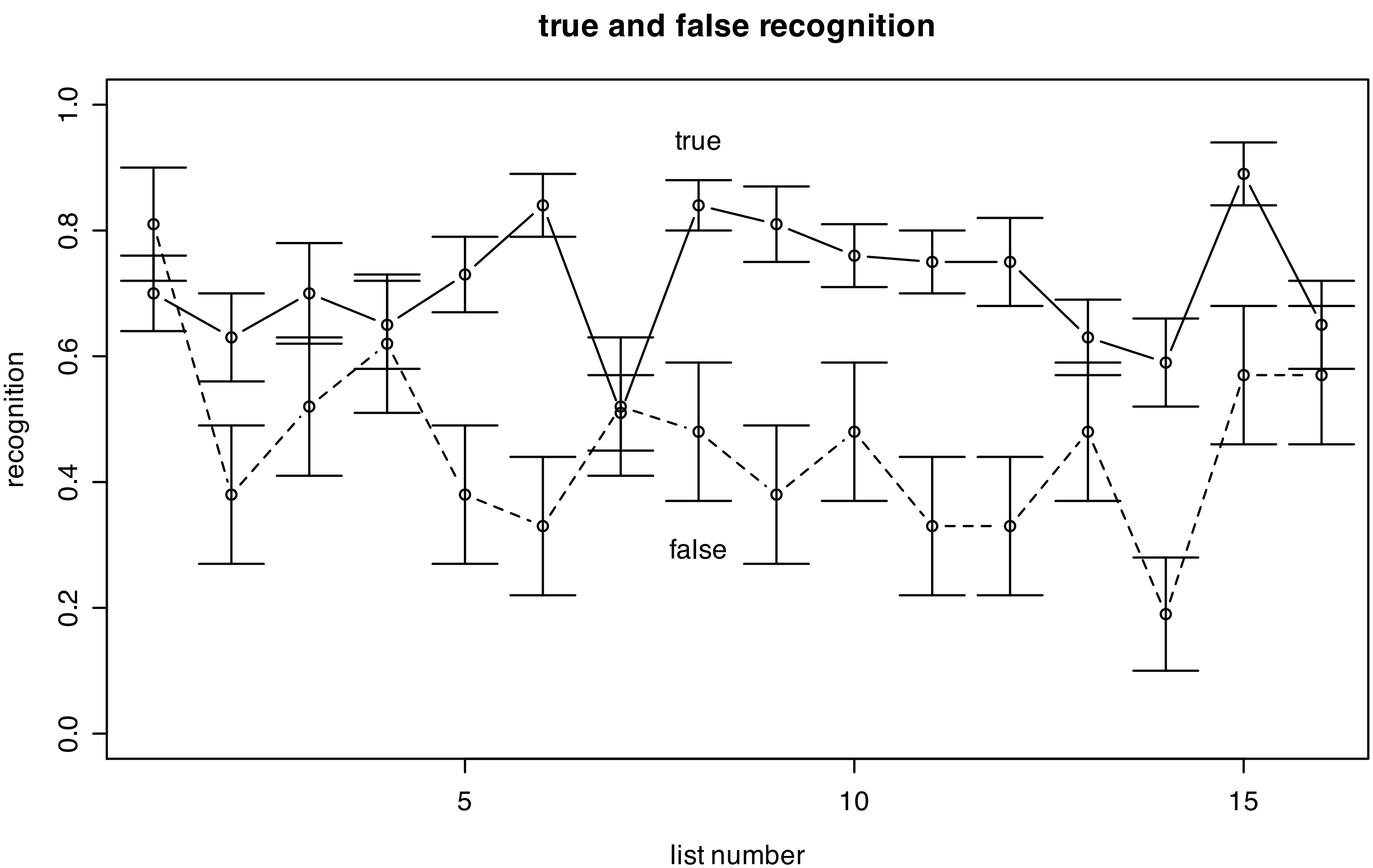
- Recall (manipulation check)
  - Is there a serial position effect?
    - Primacy
    - Recency (particularly given the instructions)
  - Did people just recall on recall tasks?
  - Do the lists differ in recall ease?
- Recognition
  - Is there a false memory effect?
  - What manipulations affect it?

P(recognition) by list order for true and false words

Central Tendencies and error

- Sample means reflect population values +/- error variability
- standard deviation of a mean (the standard error) = $s.d/\sqrt{N}$
- observed mean +/1 1 standard error includes the population value 68% of the time
- means that differ by 2.8 standard errors are unlikely to be from same population
- errors of within subject designs are more complicated to show

P(recognition) by list order for true and false words error bars show 1 standard error of the mean



Modeling the Determinants of Recognition

- Does the type of word (real versus false) make a difference
- Does recalling versus not recalling the list make a difference
- Does study time affect subsequent recognition?
- Does recall time affect subsequent recognition
- Do these variables interact with each other
  - does the relationship between x and y depend upon z?

Multiple ways to model variance

- t test compares the difference of two groups
- F-test (ANOVA) is a generalization of t to compare multiple groups
- If the independent variable is categorical, then it can be thought of in terms of groups and we can use ANOVA
- If the independent variable is continuous, then we use the linear model.
- ANOVA is a special case of linear model

Recall and Recognition Hypothesis testing

- How likely would differences of this magnitude be observed if in fact there were no effect in the population.
- Null Hypothesis Test
  - H_0 The groups do not differ in the population
  - H_1 The groups come from different populations
  - How likely are the results if H_0?
  - What is the probability of data given H_0?
  - Reject H_0 if $p <$ critical value

Alternative to hypothesis testing

- Effect size and confidence interval.
- How big is the effect and what is the expected range of the effect?

Significance testing using Analysis of Variance

- ANOVA as a generalization of t-test.
  - t-test compares the difference between two means in terms of the expected standard deviation of the mean = observed standard deviation/sqrt(N-1)
- ANOVA compares the variance of the sample means to the variance within groups
- Possible to do ANOVA for multiple comparisons (combinations of variables)

Interpretation of ANOVA

- Each anova is a comparison of two estimates of the population variance:
  - an estimate from the variance between groups and an estimate from the variance within groups.
- F is the ratio of these estimates. If the two groups are random samples from the same population, we would expect the F ratio to be 1. The more the F deviates from 1, the less likely is the hypothesis that the samples came from the same population.

Probability of recognition: 4 independent variables

- Grand mean = 0.59
- word type false real
    - 0.44 0.74
- math FALSE TRUE
    - 0.59 0.59
- study 2 3
    - 0.58 0.59
- recall 45 90
- 0.60 0.58


True and False Recognition
probability of recognition
0.0
0.2
0.4
0.6
0.8
1.0
false
real
word type

ANOVA of recognition

- Df Sum Sq Mean Sq F value Pr(>F)
- word 1 3.558 3.558 22.4693 2.618e-06 ***
- math 1 0.001 0.001 0.0094 0.9228
- study 1 0.002 0.002 0.0140 0.9058
- recall 1 0.016 0.016 0.1030 0.7483
- Residuals 659 104.360 0.158
- ---
- Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

ANOVA also can examine interactions

- Two variables have an interactive effect upon a dependent variable if the effect of the first (X1) upon the dependent (Y) depends upon the value of the second Independent Variable (X2)

Analysis of Variance main effects and interactions

		Variable 2	
		Level A	Level B
Variable 1	Level 1	1-A	1-B
	Level 2	2-A	2-B

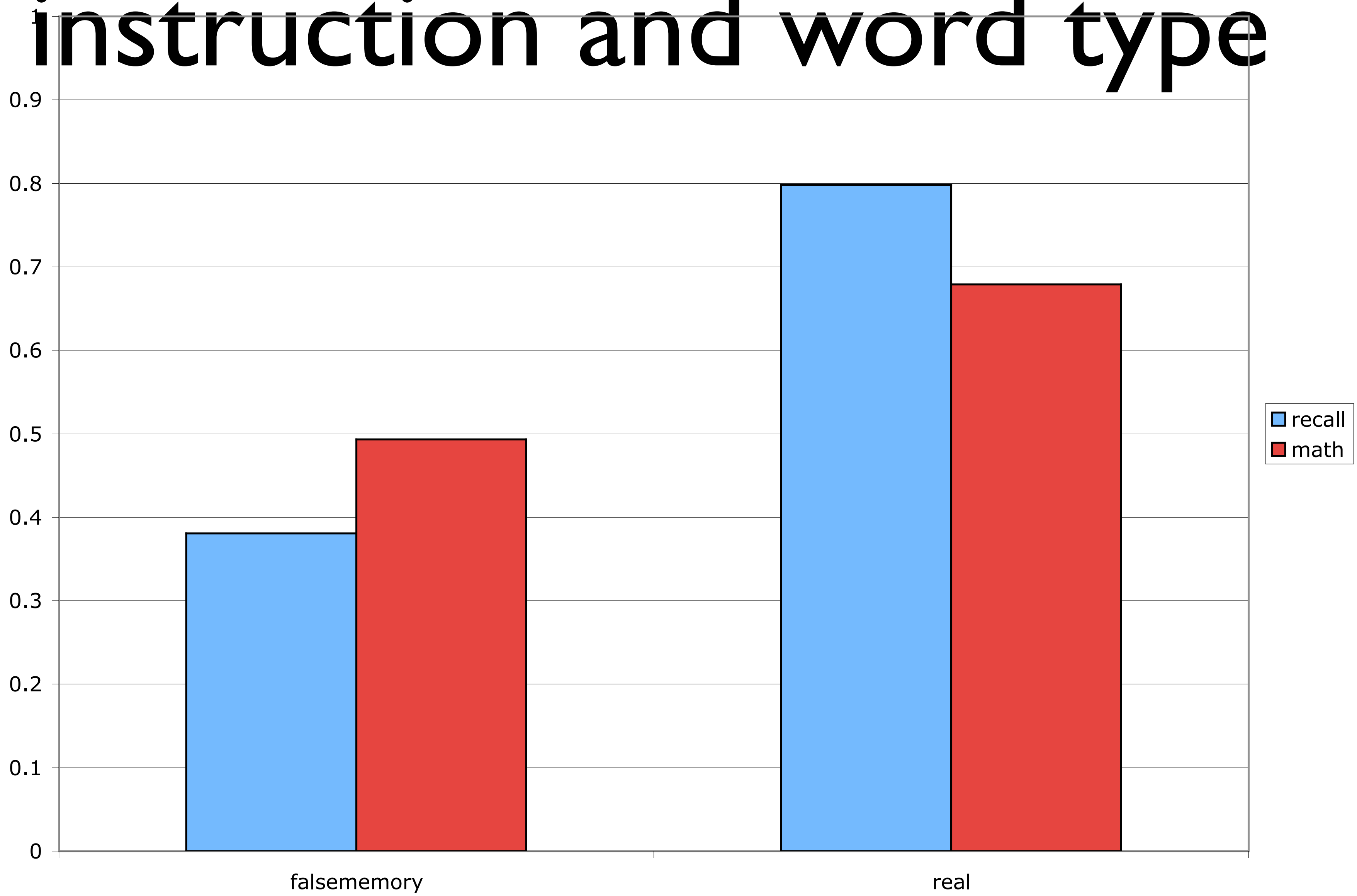
What is the effect of Variable 1?
What is the effect of Variable 2?
Does the effect of Variable 1 depend upon Variable 2

Interaction of Word type by Math/Recall

word	math FALSE	math TRUE
false	0.38	0.49
real	0.80	0.68

- (Math= False <-> Recall)

Recognition varies by instruction and word type



Two way Anova shows main effects and interaction

```
            Df  Sum Sq Mean Sq F value    Pr(>F)    
word         1   3.558   3.558 22.5588 2.497e-06 ***
math         1   0.001   0.001  0.0094 0.9226523    
word:math    1   2.263   2.263 14.3495 0.0001656 ***
Residuals 666 105.050   0.158                      
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Two way Anova shows main effects and interaction

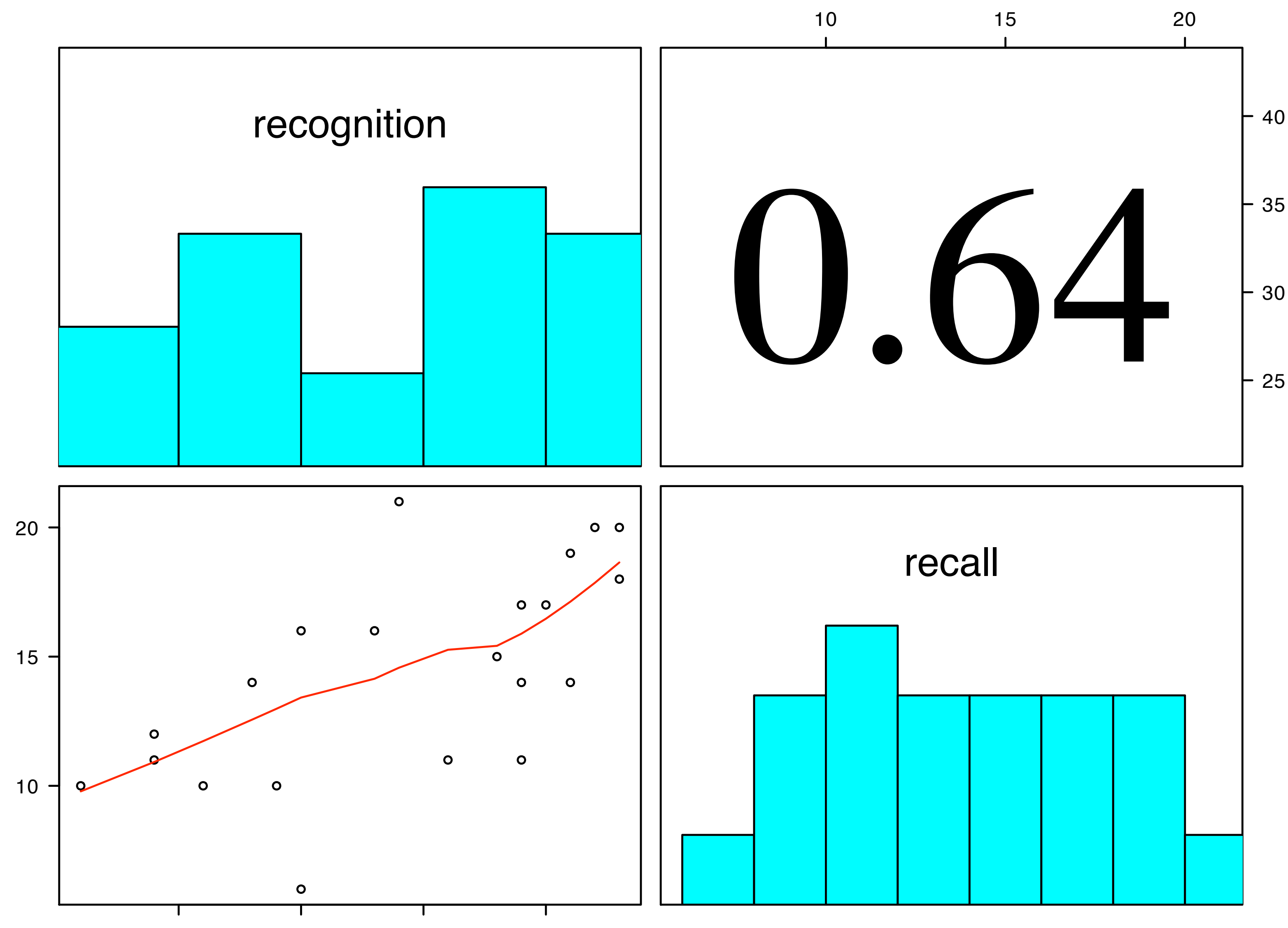
	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
word	1	3.56	3.56	22.56	0.001	***
math	1	0.00	0.00	0.01	0.92	
word:math	1	2.26	2.26	14.35	0.001	***
Residuals	666	105.05	0.16			

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

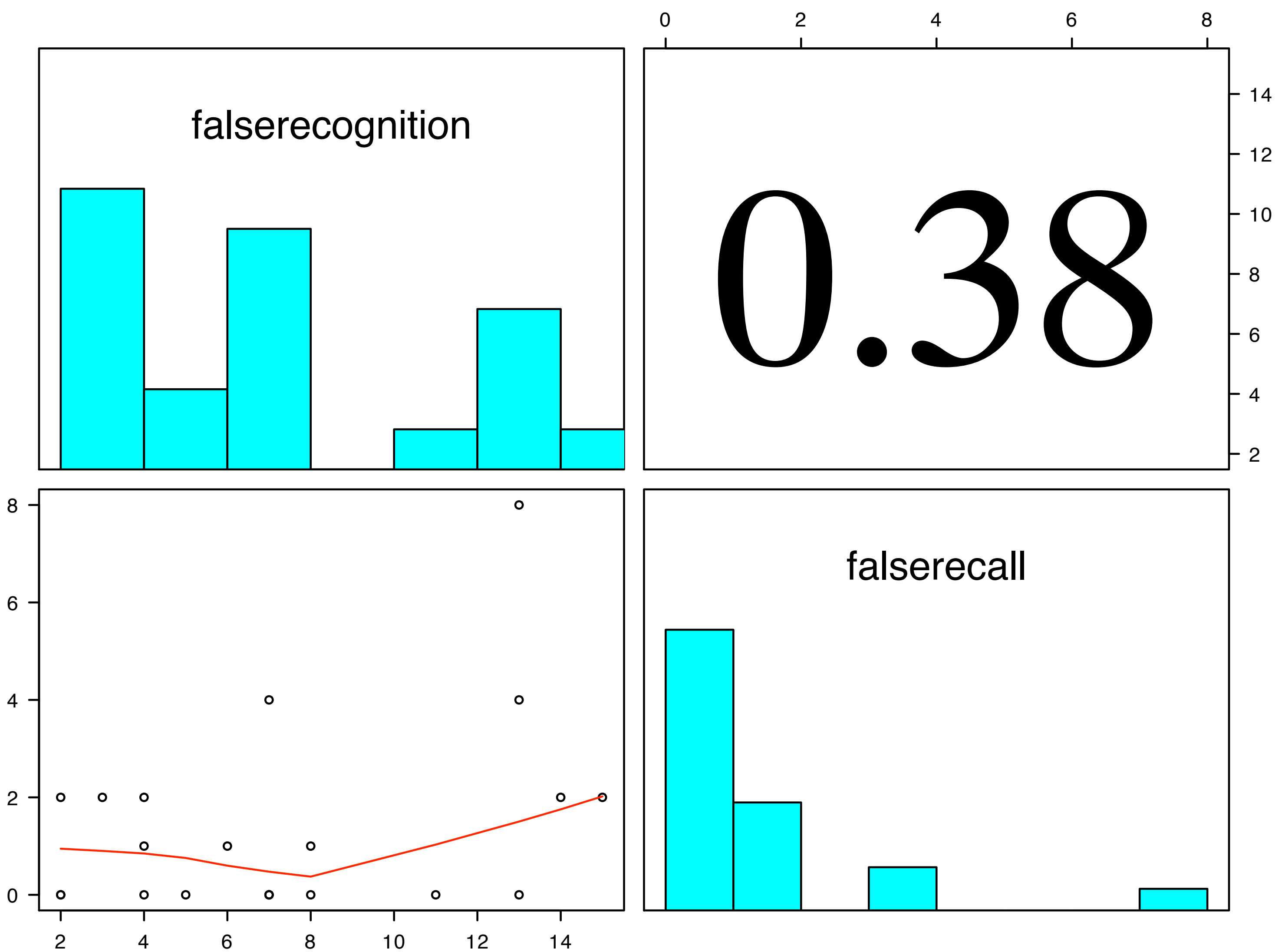
Round to 2 digits

Does recall help recognition?

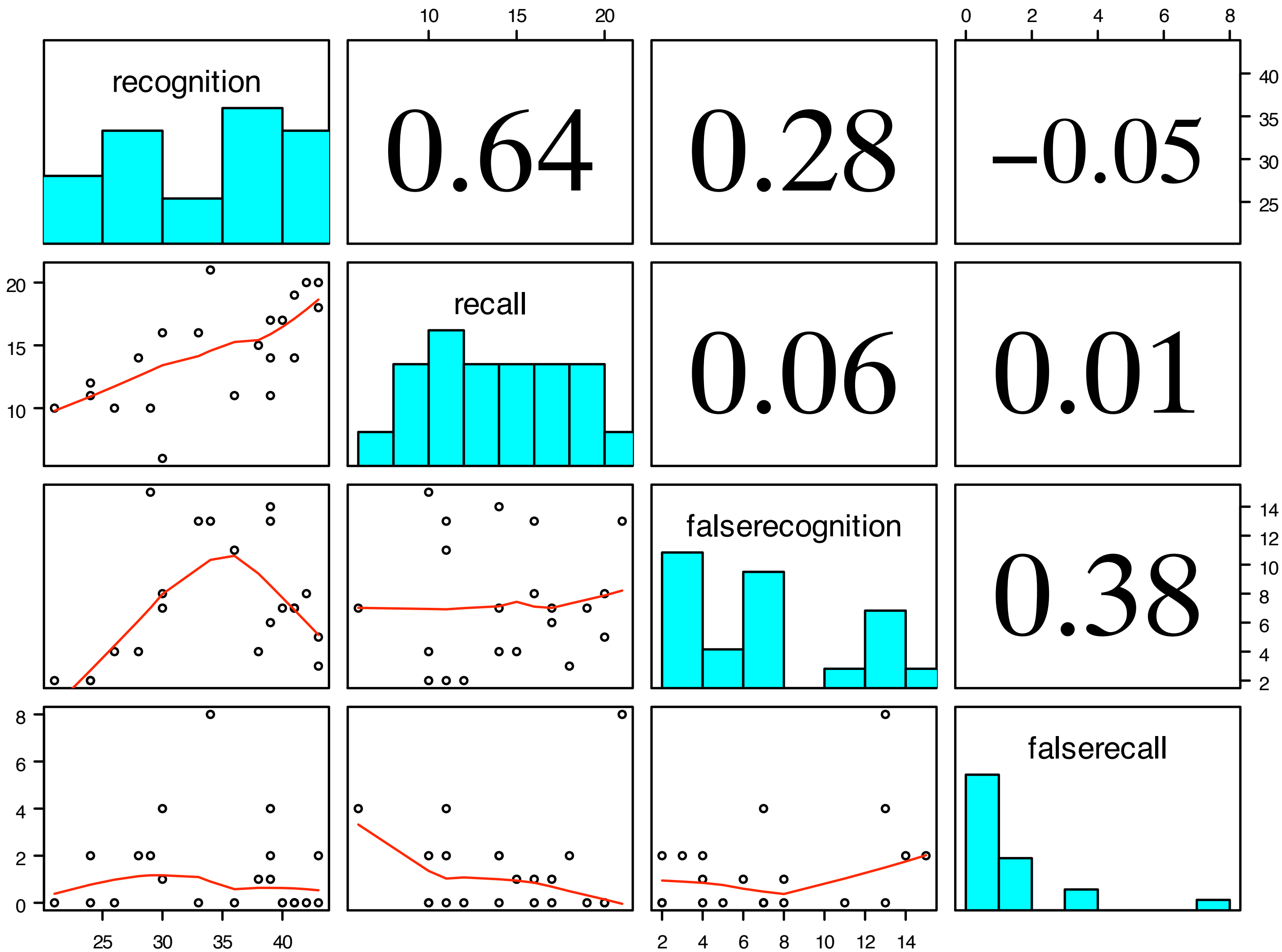
Across subjects, yes



What about for false words?



True and False Recognition



But correlations are between participants

- Perhaps some participants are better than others
  - short term memory ability
  - interests
  - language facility
- Consider the probability of recognizing given that you recalled the word (the conditional probability)

Effect of prior recall on subsequent recognition - Means

	Recalled		Not Recalled	
	Recognized	Not Recognized	Recognized	Not Recognized
Cued	0.28	0.02	0.43	0.27
False (primed)	0.07	0.02	0.39	0.52

Effect of prior recall on subsequent recognition- another perspective

	Recalled	Not Recalled
Cued	0.93	0.62
False (primed)	0.79	0.43

What do the data show?

- In words:
  - Prior recall facilitated subsequent recognition for target words and non presented but primed words
- In numbers:
  - Probability a word was identified as "old" if it had been recalled was .93 for target words and .81 for (false) primes. If not recalled, the probability of being identified as "old" was .62 for targets and .43 for false primes.

Need to convert the frequencies to conditional probabilities

- Probability of recognition | prior recall =
  - frequency of recognition and recall/ frequency of recall (with and without recognition)
- This asks the question: “if you recalled a word (either a real word or a false word) does that facilitate subsequent recognition?”
- Can examine the effect of recognition | prior recall for real and false words

Simple ANOVA
(ignoring the within subject effect)

```
aov.condit <- aov(values ~ wordtype*recalltype, data=condit.df)
```

```
                     Df Sum Sq Mean Sq F value    Pr(>F)
wordtype              1 0.81   0.81 15.98       0.0001566 ***
recalltype            1 1.86   1.86 36.79       5.996e-08 ***
wordtype:recalltype   1 0.02   0.02  0.34       0.5607753
Residuals            70 3.54   0.05
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Interpreting conditional recognition

- Prior recall facilitated subsequent recognition
- False words were recognized less than real words
- But, the effect of prior recall did not differ between false and real words.
  - Does this imply that the processes are the same? No, but
  - We can not show that the processes differ

Conclusions

- Big picture
  - Possible to show false memory, particularly in a recognition task
- Smaller picture
  - variables that affect false recognition
- Take home message:
  - What does this all mean