What we did

1. ANOVA with pooled variances
2. Sample size is too small to check assumptions in residuals; instead, compare results against permutation tests
3. Inspected images to arbitrarily chose pairs or groups within the same “batch”
4. **Bonus:** use ANOVA on the entire 120 groups for a singel voxel, plotted top 10 “hits”

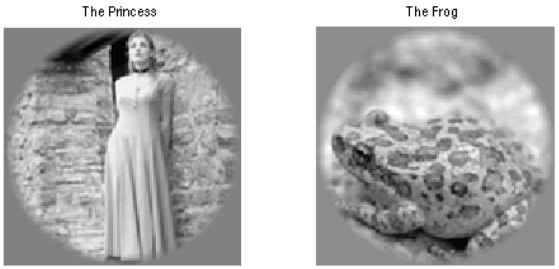
Pairwise test

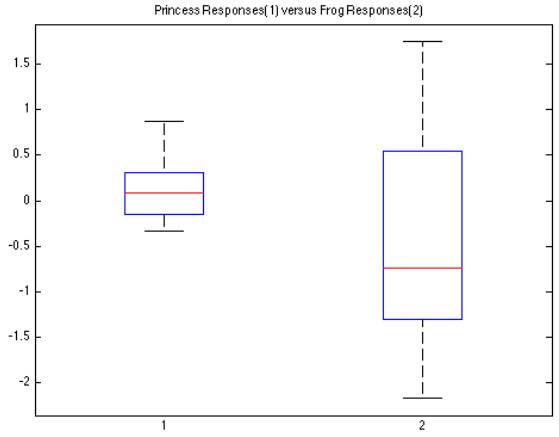
p=0.08 (t-test)

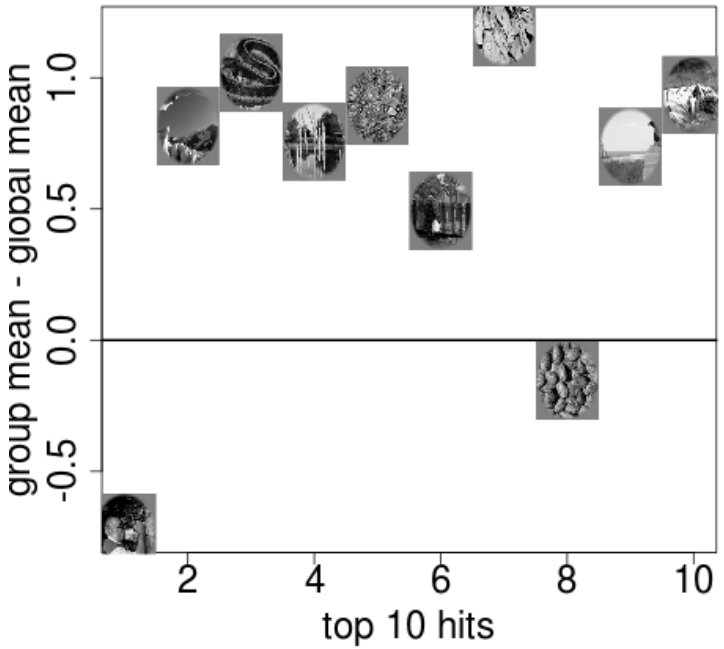
p=0.10 (permutation)

mean=0.62

CI=???







ANOVA on 120

top 10 p-values range from 1e-18 to 1e-8

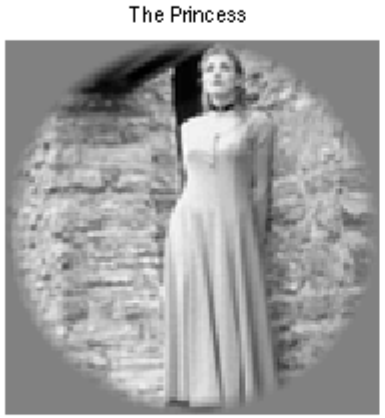
F-statistic: $p < 2.2e-16$

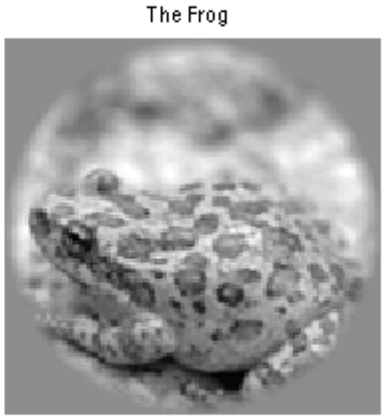
Results

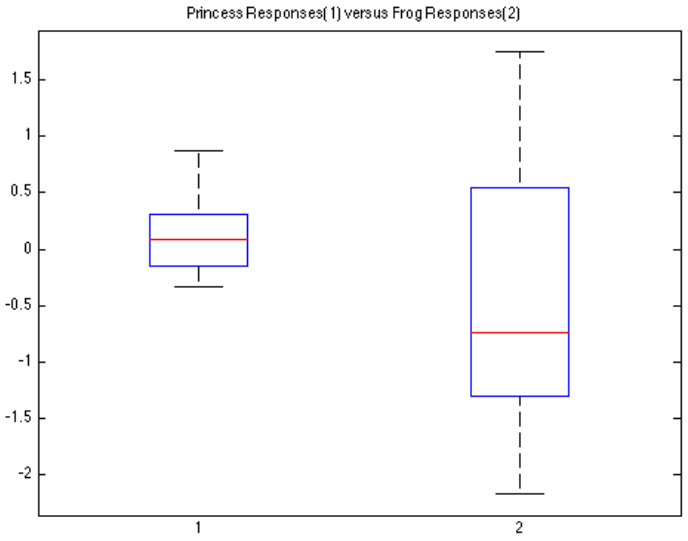
p=0.08

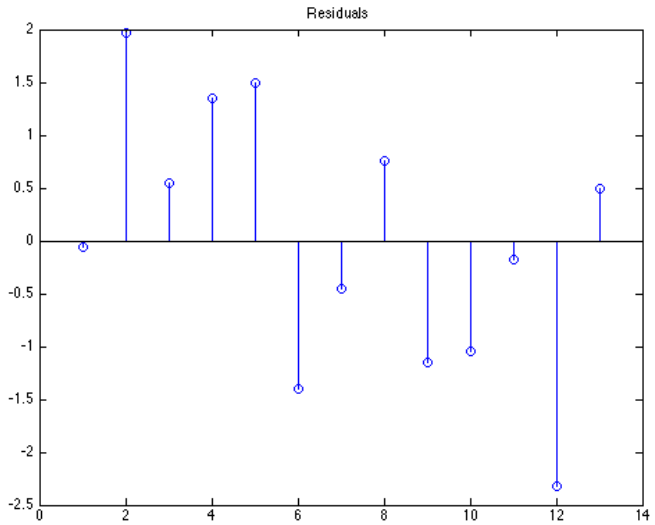
mean=0.62

CI=(-0.093,1.34)









Model setup

Expectation is linear, so
E[A-B]=E[A]-E[B]

We assumed that the responses to the images (or groups of images) were independent, Gaussian random variables. In this case, the variances simply add
standard error(A-B) = (var(A) + var(B))/N
with N degrees of freedom

Therefore, calculating the p-value comes down to the probability that a difference more extreme than the observed difference between the sample mean responses is drawn from a student's T distribution with mean 0 and standard deviation given by the standard error above.