Thoughts about displaying data

William Revelle, Joshua Wilt, David Condon and Ashley Brown-Riddell
PMC lab
Northwestern University

Abstract

We have known since Cattell's discussion of the Data Box that we may organize the raw data collected in psychology as either a two dimensional array of M observations or tests (rows) on N persons (columns), or, if using short or long term longitudinal designs a three dimensional array with the third dimension being O observations across occasions Cattell (1946). How best to display and conceptualize these data remains a challenging problem. The traditional approach is to consider how tests correlate over people (Cattell's R analysis) or how people correlate over tests (Q analysis), although the ability to do multilevel modeling and growth curve analysis allows for examining the structure of individual people as they vary on tests over occasions. We will limit ourselves to a consideration of how to display R data (N people over M tests).

The structure of emotion words

Consider the self reported mood data on 75 adjectives (the Motivational State Questionnaire or MSQ) collected from 3,896 participants over nine years at the Personality, Motivation and Cognition laboratory as part of a series of studies examining the effects of personality and situational factors on motivational state and subsequent cognitive performance (Revelle & Anderson, 1997; Rafaeli & Revelle, 2006). In each of 38 studies, prior to any manipulation of motivational state, participants signed a consent form and filled out the MSQ. (The procedures of the individual studies are irrelevant to this data set and

---

contact: William Revelle revelle@northwestern.edu
Draft version of

could not affect the responses to the MSQ, since this instrument was completed before any further instructions or tasks). These data are publicly available as the `msq` data set in the *psych* package (Revelle, 2011) in the open source statistical language R (R Development Core Team, 2011). Because not all items were given to all subjects, and one item (quiescent) has a very low communality with the rest of the words, we analyze 69 of the mood terms.

Perhaps the clearest demonstration of structure is a "heat map" representation of the correlations, where positive correlations are represented by blue, negative correlations by red, and magnitude of correlations by saturation (Figure 1).

```
> data(msq)
> #irt.msq <- irt.fa(msq[1:70],2)   #%already done
> msqp <- polar(irt.msq$fa)
> neword <- msqp[,1]
> msqr <- irt.msq$rho[neword,neword]
> msqr <- msqr[-16,-16]  #no communality
> f2 <- fa(msqr,2)
> summary(f2)

Factor analysis with Call: fa(r = msqr, nfactors = 2)

Test of the hypothesis that 2 factors are sufficient.
The degrees of freedom for the model is 2209  and the objective function was  28.36

The root mean square of the residuals (RMSA) is  0.07
The df corrected root mean square of the residuals is  0.1

 With factor correlations of
     MR1  MR2
MR1  1.0 -0.1
MR2 -0.1  1.0
```

A conventional representation using factor analysis

A conventional way to represent the complexity of the data is to project the 69 dimensions (the individual variables) onto two dimensions (or factors).

The adequacy of the factor model may be evaluated in how well it accounts for the observed correlations. This may be seen graphically by a *heat map* of the residual correlations (Figure 3). It is clear that the traditional two dimensional representation of these affect terms fails to account for the correlations between the low negative affect terms (serene, calm, etc.), and that a three factor solution might be more applicable. When this is done, the residuals are much smaller (Figure 4).

Graph theoretic descriptions

When we plot the correlations using the recommended package, *qgraph* we do not see such a clear bifactor solution, nor is it clear how to find that the the two factor solution is inadequate.

```
> pdf( 'corplot.pdf' )
> cor.plot(msqr,cex.axis=.5,main='A cor.plot of 69 emotion words')
> dev.off()

quartz
     2
```

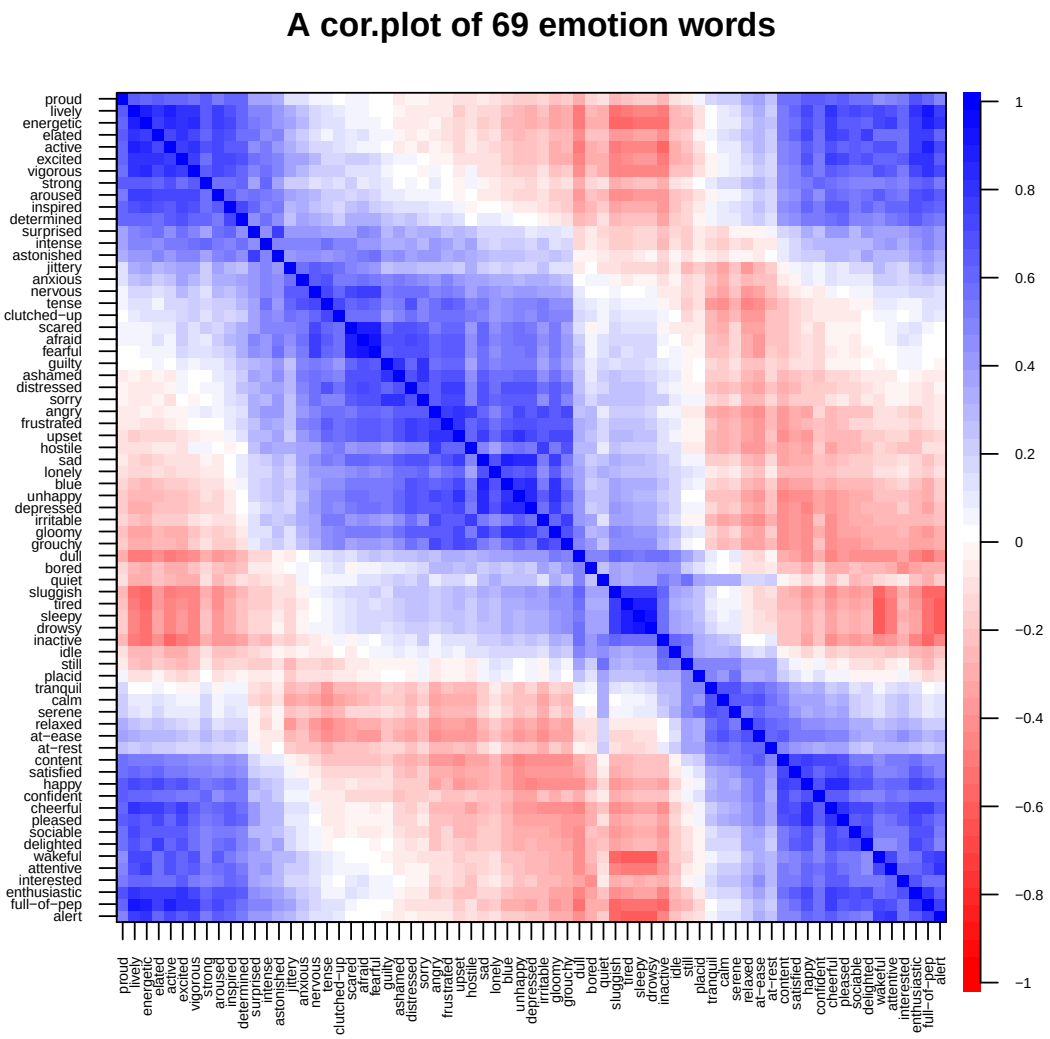

Figure 1. A "heat map" display of the correlations between 69 mood adjectives. The pattern represents the circumplex structure of the data, with high positive correlations along the main diagonal, fading to zero and then negative and then back to positive as the items become progressively further apart.

```
> factor.plot(f2,title='Factor analysis of 69 emotion words')
> f3 <- fa(msqr,3)
```

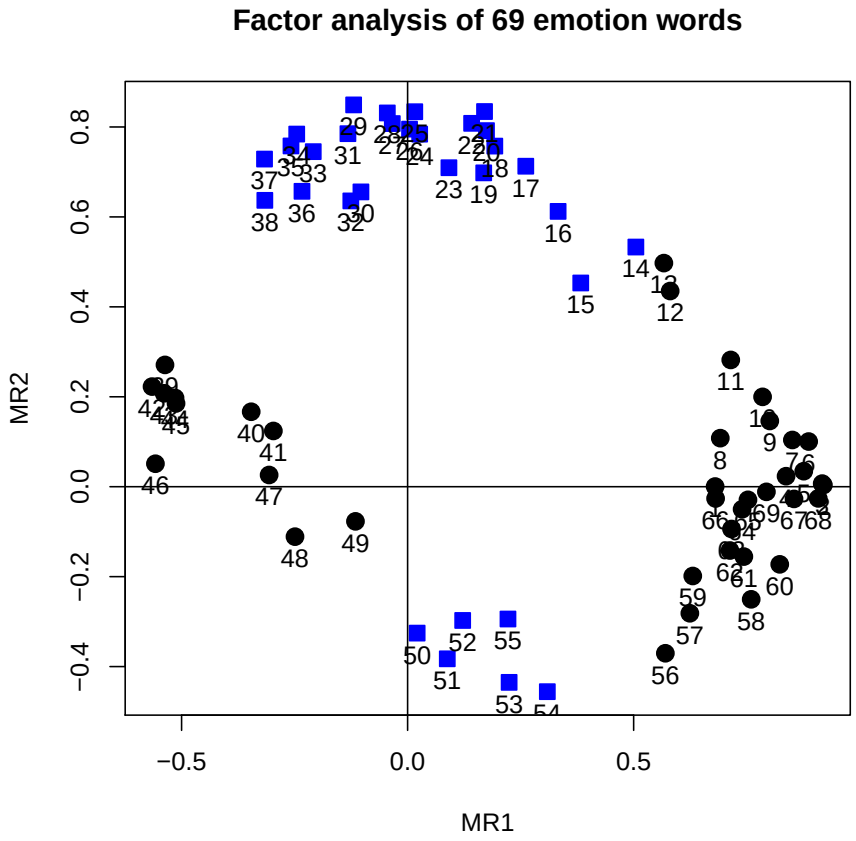

Figure 2. A two dimensional factor plot of the loadings of 69 msq items. The item numbers correspond to the item orders found in Figure 1).

```
> pdf('residplot.pdf')
> cor.plot(resid(f2),cex.axis=.5,main='Residual correlations after extracting 2 factors')
> dev.off()

quartz
     2
```

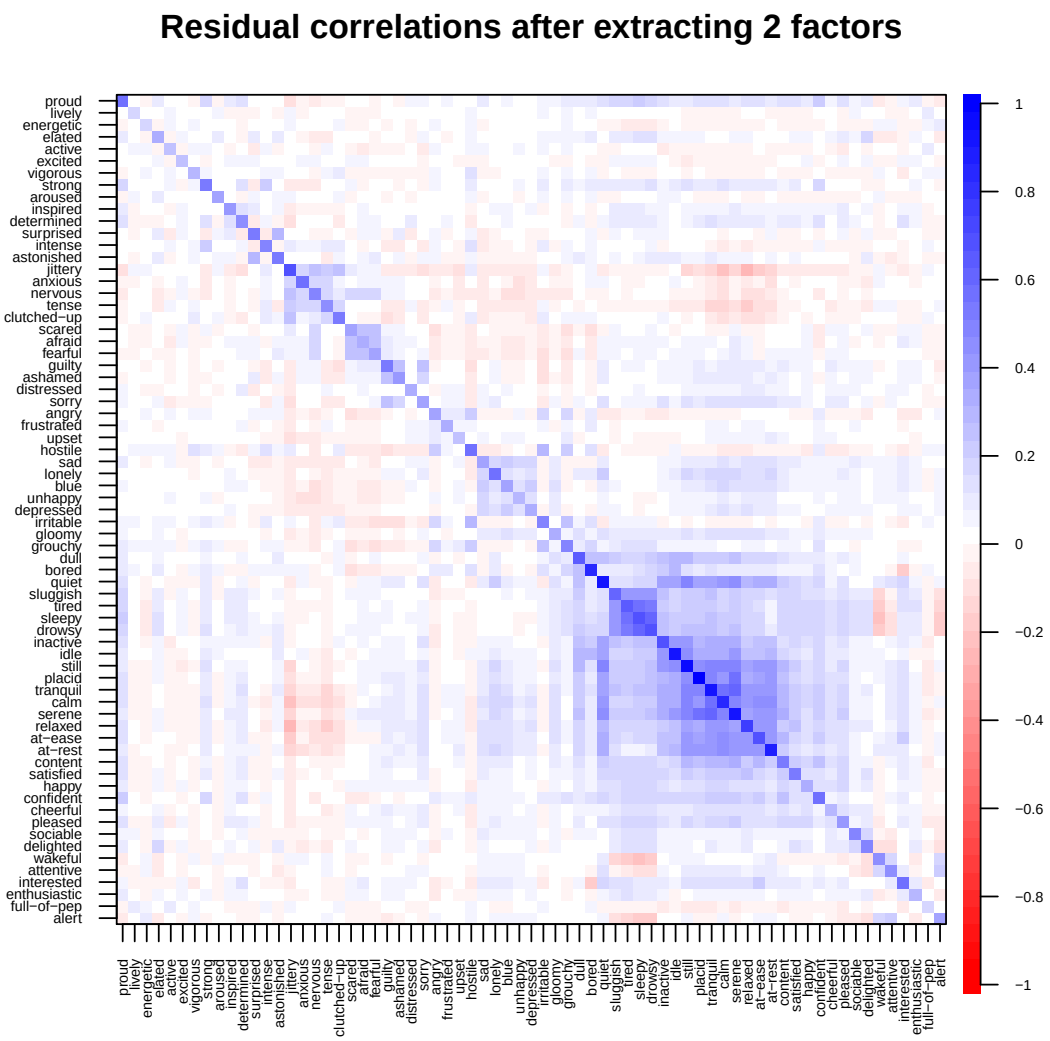

Figure 3. The residual correlations suggest that the low negative affect terms are not well fit by a two dimensional model.

```
> pdf('residplot3.pdf')
> cor.plot(resid(f3),cex.axis=.5,main='Residual correlations after extracting 3 factors')
> dev.off()

quartz
     2
```

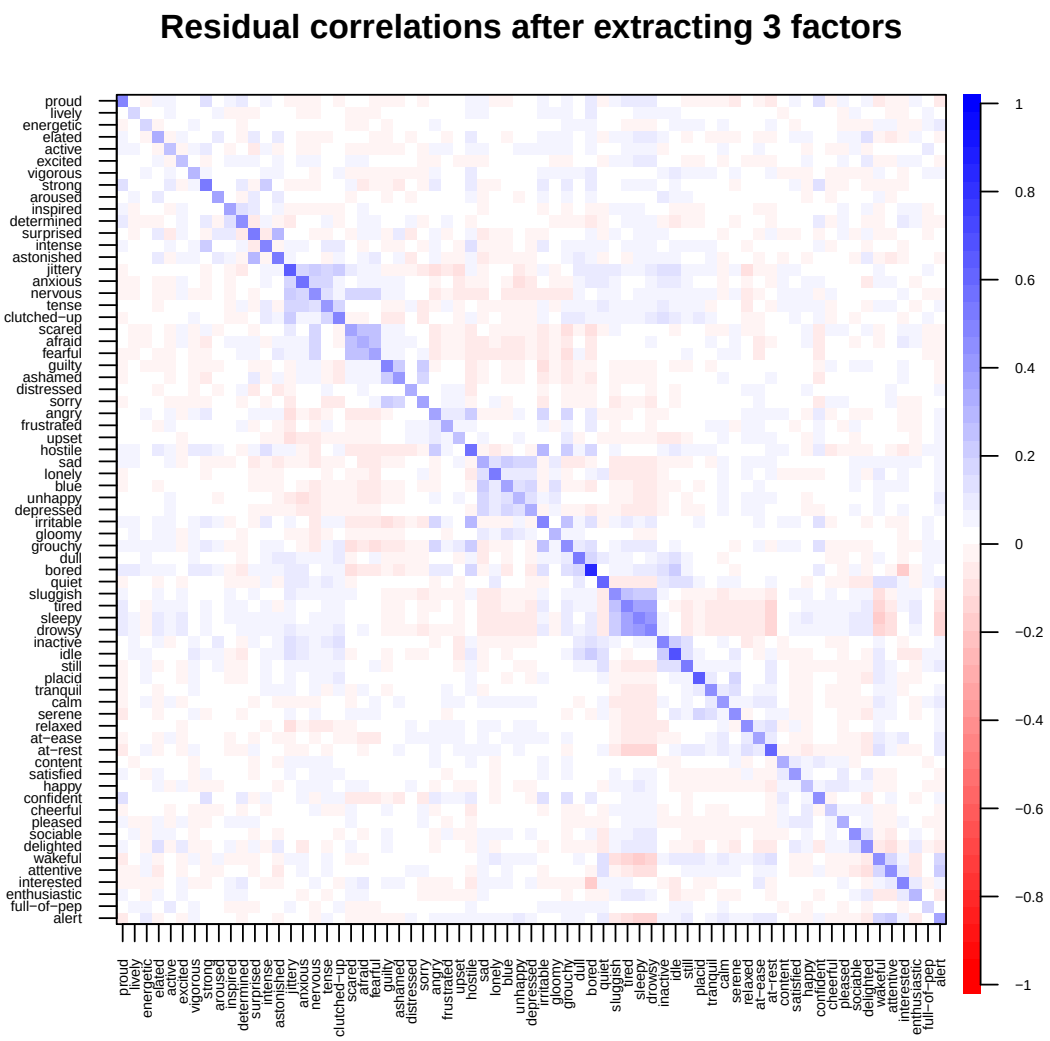

Figure 4. The residual correlations from a three factor solution are quite small and suggest that the negative affect dimension should be considered in terms of independent dimensions of negative affect and calmness/serenity.

```
> qgraph(msqr,layout='spring')
```

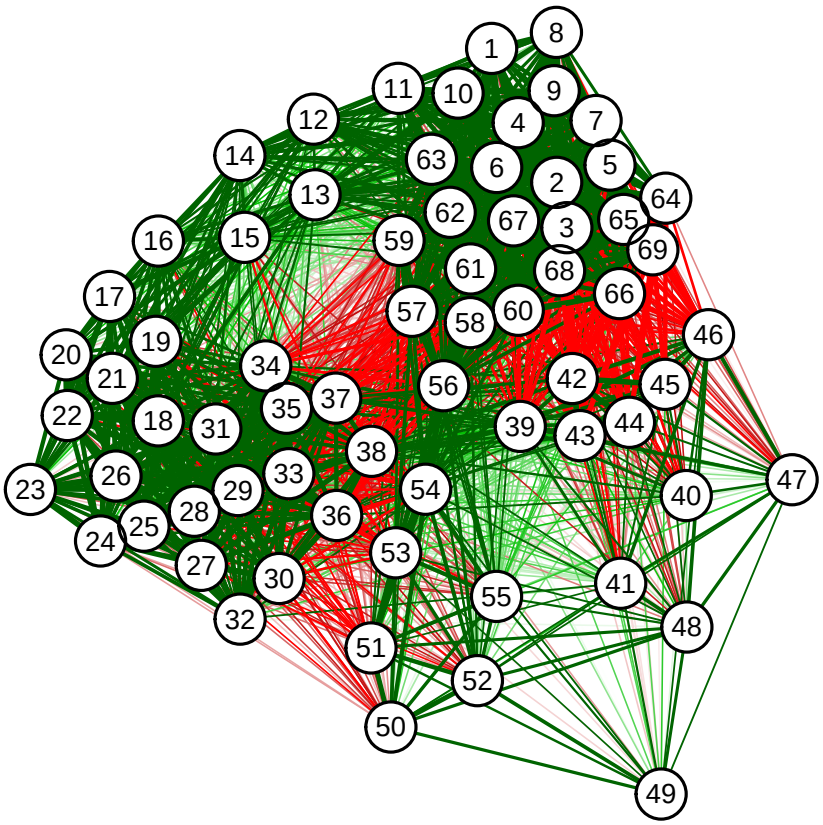

Figure 5. The *qgraph* representation of the msq correlation matrix does not capture the bipolar nature of the affective words.

```
> op <- par(mfrow=c(2,2))
>  qgraph(msqr,layout="spring",minimum=.6)
>  qgraph(msqr,layout="spring",minimum=.5)
>  qgraph(msqr,layout="spring",minimum=.4)
>  qgraph(msqr,layout="spring",minimum=.3)
>  op <- par(mfrow=c(1,1))
```

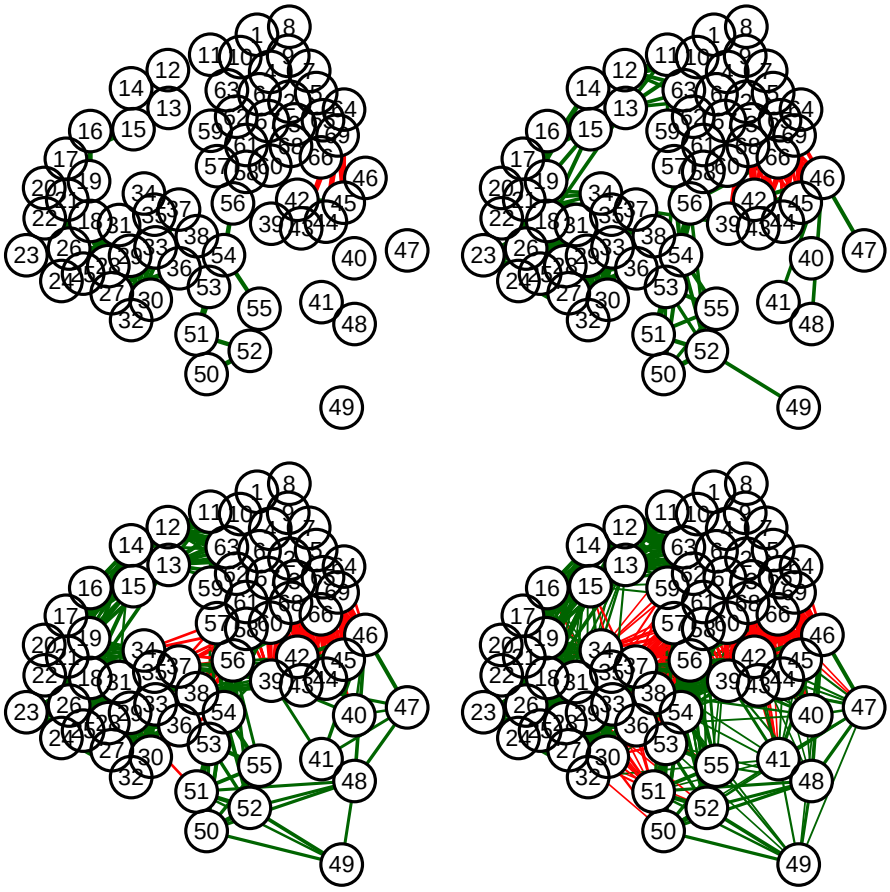

Figure 6. Changing the minimum value to plot allows for some structure to be seen.

Factors versus graph theory

When doing the factor analysis, the model of the data is descriptive, but does allow for a test of the level of fit. It is unclear how to determine whether one beautiful graph fits or does not fit the data.

References

Cattell, R. B. (1946). Personality structure and measurement. I. The operational determination of trait unities. *British Journal of Psychology*, *36*, 88-102.

R Development Core Team. (2011). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Available from `http://www.R-project.org/` (ISBN 3-900051-07-0)

Rafaeli, E., & Revelle, W. (2006). A premature consensus: Are happiness and sadness truly opposite affects? *Motivation and Emotion*, *30*(1), 1-12.

Revelle, W. (2011). psych: Procedures for personality and psychological research [Computer software manual]. Evanston. Available from `http://personality-project.org/r` (R package version 1.1.12)

Revelle, W., & Anderson, K. J. (1997). *Personality, motivation and cognitive performance* (Final report to the Army Research Institute on contract MDA 903-93-K-0008). Northwestern University:.