Grizzly Factor Analysis and PCA

Elizabeth Hiroyasu

January 25, 2018

Factor analysis conducted to assess how people think about grizzly bear reintroduction and what types of variables those attitudes map onto.

We are specifically interested in people's attitudes toward the reintroduction of grizzly bears. Respondents were asked to respond with how much they agreed or disagreed on a 7-point likert scale with the following questions: 1. People have a responsibility to ensure the survival of grizzly bears. 2. Grizzly bear reintroduction would help make California forests healthier. 3. Grizzly bear reintroduction would pose a threat to my safety. *4. Grizzly bear reintroduction would pose a threat to my livelihood.* 5. Grizzly bear reintroduction would benefit the California economy by increasing tourism. 6. Grizzly bear reintroduction would lead to an increased role for the federal government. 7. Grizzly bear reintroduction would threaten property rights on private lands. 8. Grizzly bear reintroduction would benefit other species. 9. Grizzly bear reintroduction would help prevent their extinction. 10. Grizzly bear reintroduction would reduce local control over public lands. 11. Grizzly bear reintroduction would negatively impact ranchers. 12. Grizzly bear reintroduction would benefit outdoor recreation. 13. Grizzly bear reintroduction would harm agricultural producers. 14. Grizzly bear reintroduction would benefit urban residents. 15. Grizzly bear reintroduction would benefit rural residents.

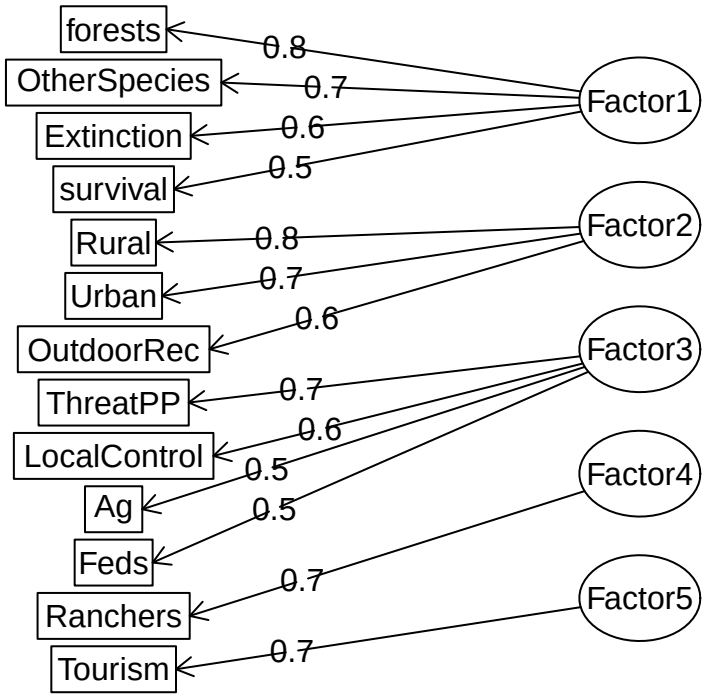
*Questions 3 and 4 were dropped out of this analysis and analyzed separately, because they are threats about personal impacts of reintroducing grizzly bears.

Pooled Factor Analysis

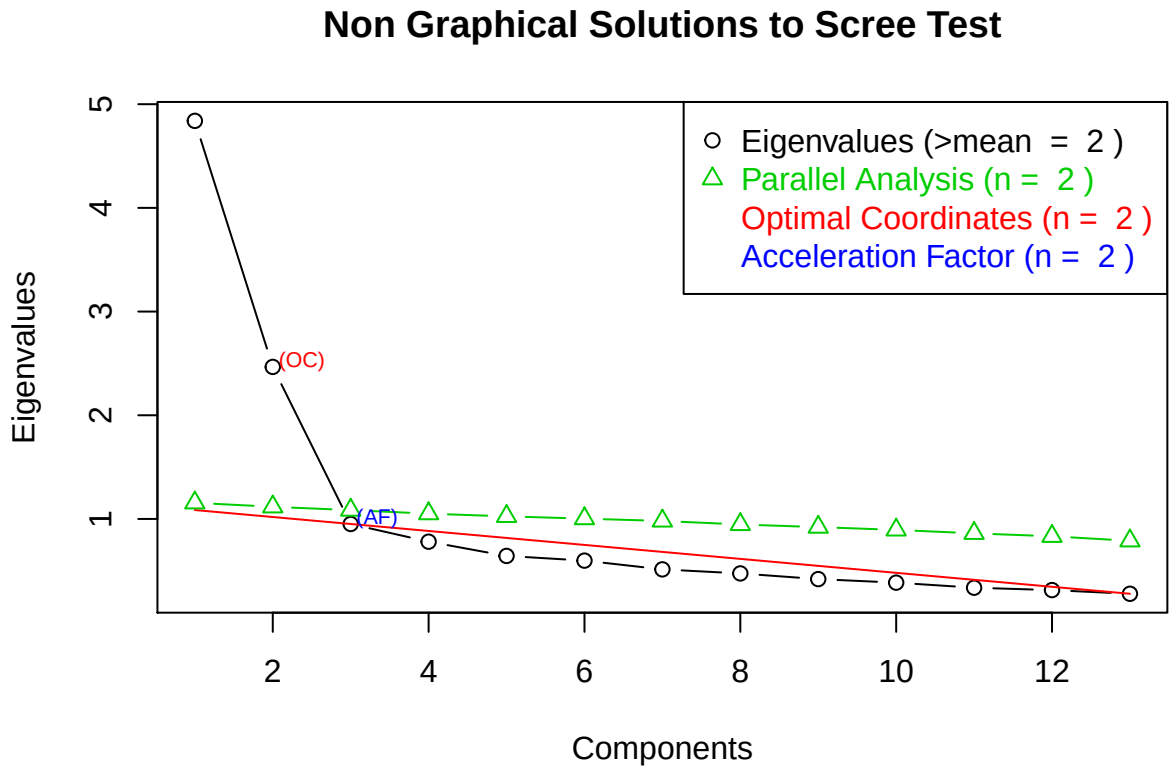
The factanal command uses the maximum likelihood to identify the loadings of different components. We used an a priori guess of five factors to load on.

```
##
## Call:
## factanal(x = grizz_factor, factors = 5, scores = "Bartlett")
##
## Uniquenesses:
##     survival      forests      Tourism         Feds     ThreatPP
##        0.658        0.282        0.136        0.762        0.356
## OtherSpecies   Extinction LocalControl     Ranchers   OutdoorRec
##        0.348        0.573        0.607        0.204        0.401
##           Ag        Urban        Rural
##        0.469        0.420        0.216
##
## Loadings:
##              Factor1 Factor2 Factor3 Factor4 Factor5
## survival      0.545   0.167          -0.103
## forests       0.768   0.324                   0.112
## Tourism       0.442   0.418                   0.697
## Feds                          0.480
## ThreatPP     -0.220           0.722   0.255
## OtherSpecies  0.704   0.384
## Extinction    0.611   0.175          -0.112   0.103
## LocalControl          0.122   0.599   0.125
## Ranchers     -0.212  -0.201   0.498   0.676
```

```
## OutdoorRec    0.442   0.588                    0.216
## Ag           -0.212           0.507   0.476
## Urban         0.336   0.664   0.128
## Rural         0.329   0.812
##
##                 Factor1 Factor2 Factor3 Factor4 Factor5
## SS loadings       2.513   1.997   1.644   0.823   0.593
## Proportion Var    0.193   0.154   0.126   0.063   0.046
## Cumulative Var    0.193   0.347   0.473   0.537   0.582
##
## Test of the hypothesis that 5 factors are sufficient.
## The chi square statistic is 30.18 on 23 degrees of freedom.
## The p-value is 0.144
```



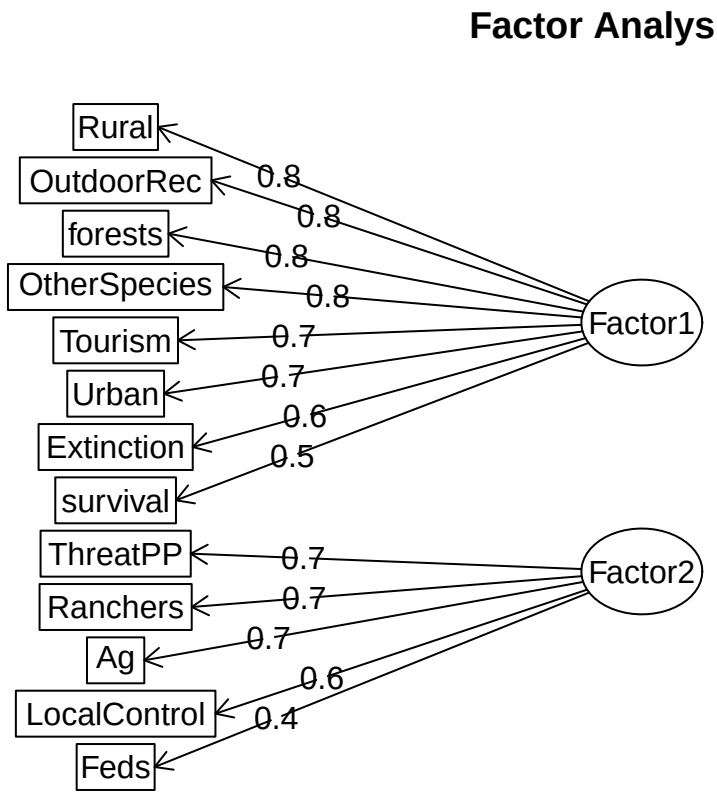
Suppose we want to determine ahead of time the number of factors to extract:



This suggests that the 15 questions collapse into two factors. We can then rerun the factor analysis, specifying two factors.

```
##
## Call:
## factanal(x = grizz_factor, factors = 2, scores = "Bartlett")
##
## Uniquenesses:
##     survival      forests      Tourism         Feds     ThreatPP
##        0.718        0.387        0.477        0.846        0.399
## OtherSpecies   Extinction LocalControl     Ranchers   OutdoorRec
##        0.419        0.654        0.645        0.374        0.401
##           Ag        Urban        Rural
##        0.466        0.480        0.399
##
## Loadings:
##              Factor1 Factor2
## survival      0.524
## forests       0.771  -0.135
## Tourism       0.723
## Feds                  0.380
## ThreatPP     -0.201   0.749
## OtherSpecies  0.754  -0.111
## Extinction    0.574  -0.128
```

```
## LocalControl  0.143   0.579
## Ranchers     -0.341   0.714
## OutdoorRec    0.772
## Ag           -0.176   0.709
## Urban         0.708   0.140
## Rural         0.774
##
##                 Factor1 Factor2
## SS loadings       4.205   2.130
## Proportion Var    0.323   0.164
## Cumulative Var    0.323   0.487
##
## Test of the hypothesis that 2 factors are sufficient.
## The chi square statistic is 411.06 on 53 degrees of freedom.
## The p-value is 7.56e-57
```

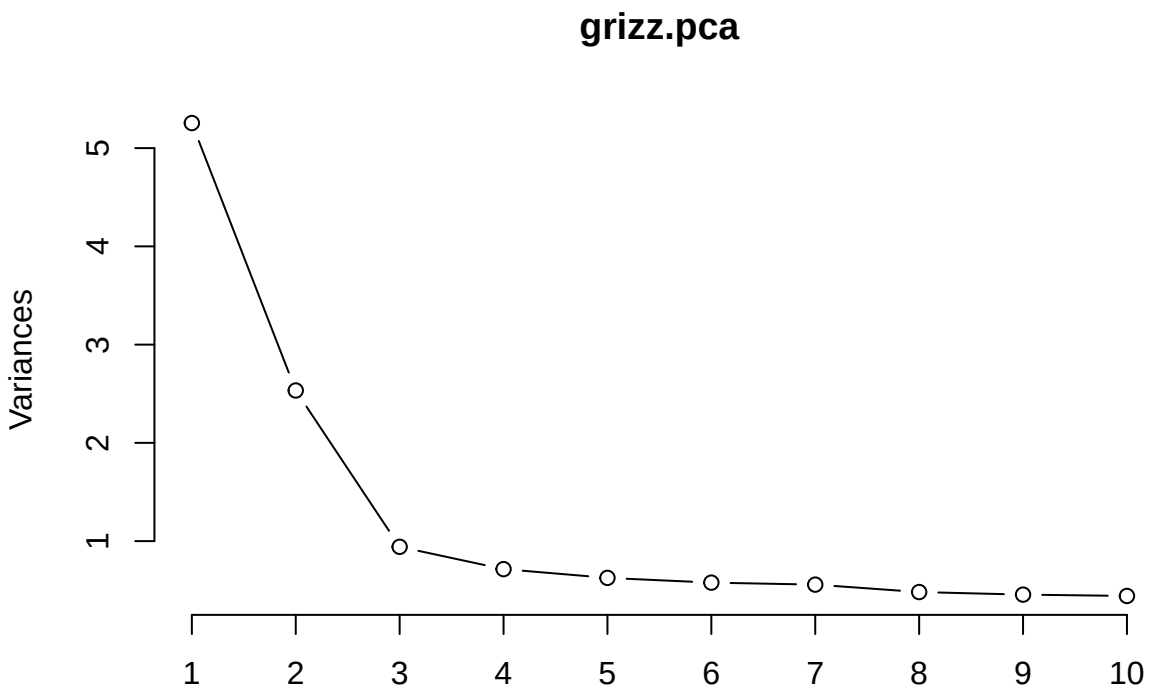


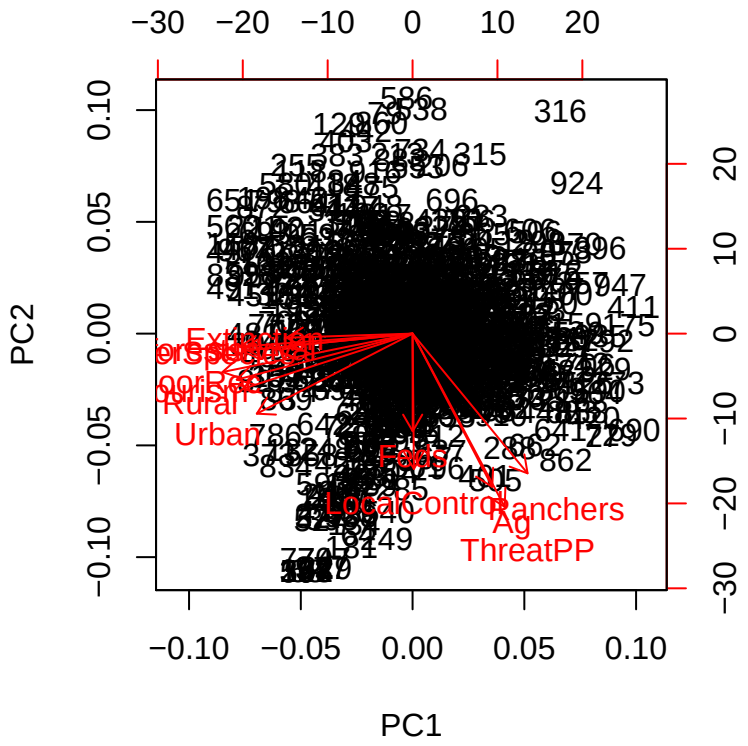
Pooled PCA:

note:prcomp is preferred to princomp The pooled PCA also suggests that the questions collapse onto two main factors

```
## Importance of components:
##                           PC1    PC2     PC3     PC4     PC5    PC6     PC7
## Standard deviation     2.292 1.5916 0.97055 0.84541 0.79100 0.7597 0.74630
## Proportion of Variance 0.387 0.1865 0.06936 0.05263 0.04607 0.0425 0.04101
## Cumulative Proportion  0.387 0.5735 0.64286 0.69549 0.74156 0.7841 0.82508
```

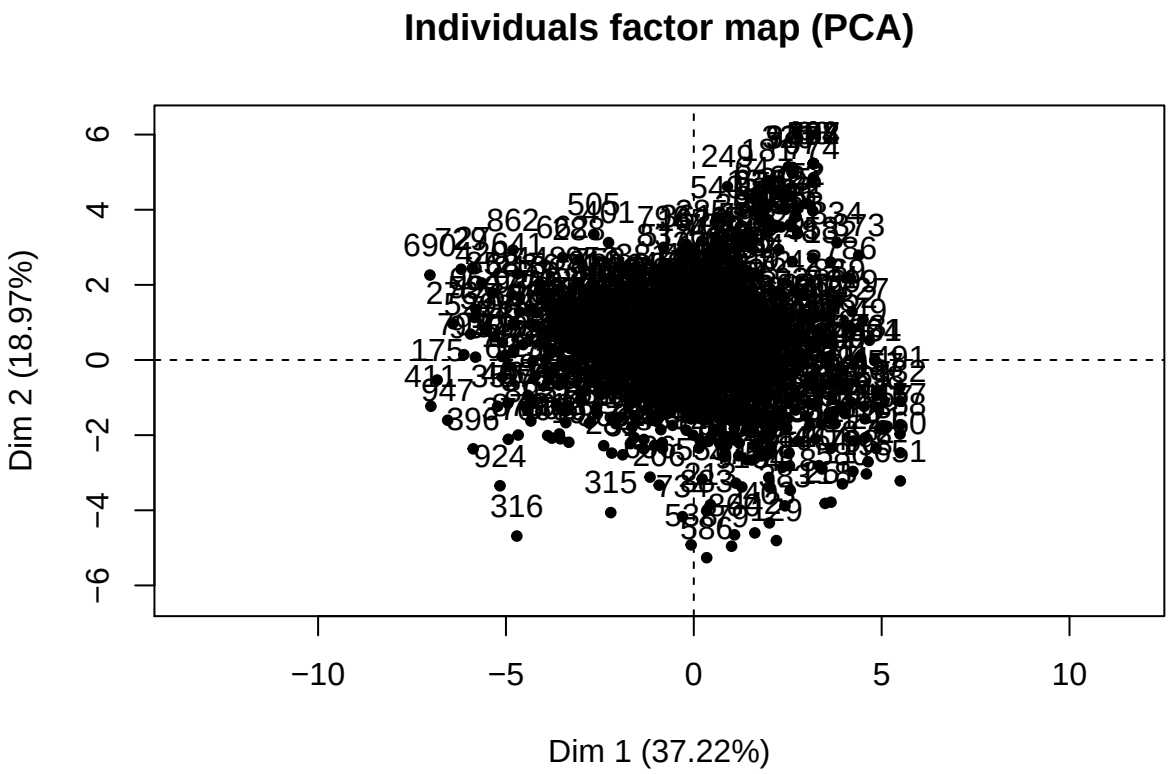
```
##                             PC8     PC9    PC10    PC11    PC12    PC13
## Standard deviation     0.69385 0.67475 0.66460 0.60745 0.57767 0.54253
## Proportion of Variance 0.03545 0.03353 0.03252 0.02717 0.02457 0.02167
## Cumulative Proportion  0.86053 0.89406 0.92658 0.95375 0.97833 1.00000
```

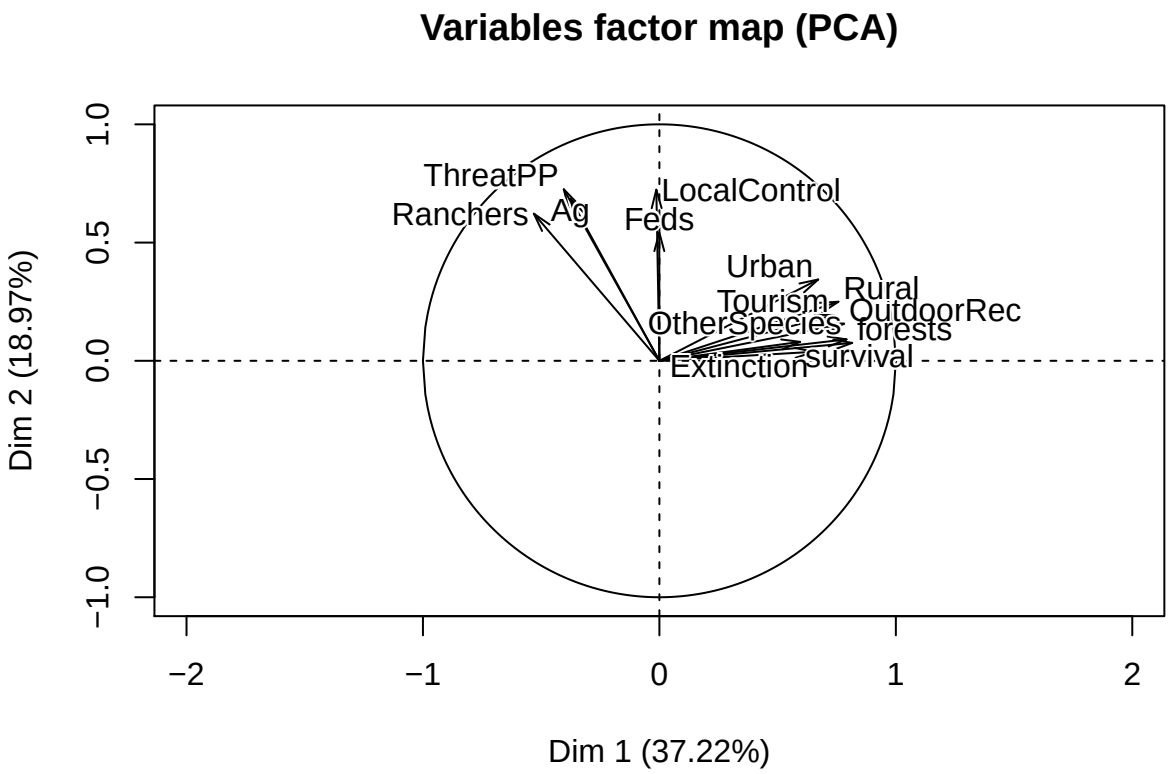




Using the FactoMineR package for more exploratory principal components analysis

```
grizz.pca2<-PCA(grizz_factor)
```


Individuals factor map (PCA)
Dim 2 (18.97%)
Dim 1 (37.22%)
6
4
2
0
−2
−4
−6
−10
−5
0
5
10
316
924
315
586




```
scores<-grizz.pca2$ind$coord
```

Pooled NMDS

currently no convergence Probably because there isn't a variable that binds the questions together. We should have N objects measured on p numeric variables. In this case, we just have p numeric variables, without the objects.