



face2gene:

## Latent Profile Analysis within Diagnosis

Virgilio Gonzenbach, Kelly Clark, Russell T. Shinohara

[vgonzenb@pennmedicine.upenn.edu](mailto:vgonzenb@pennmedicine.upenn.edu)

PennSIVE

2022-02-23

# Latent Profile Analysis

Latent Profile Analysis, also known as Gaussian Mixture Models, is a model-based clustering approach aimed at finding distinct "profiles", or clusters, based on continuous data.

The present analysis consists of the following three cases according to sample and pre-processing method selected:

- 22q: PCA was performed on gestalt scores from 22q11.2 deletion sample.
- PS: PCA was performed on gestalt scores from Psychosis Spectrum (PS) groups.
- projPS: PS gestalt scores were projected onto 22q PCA results.

Summary of procedure:

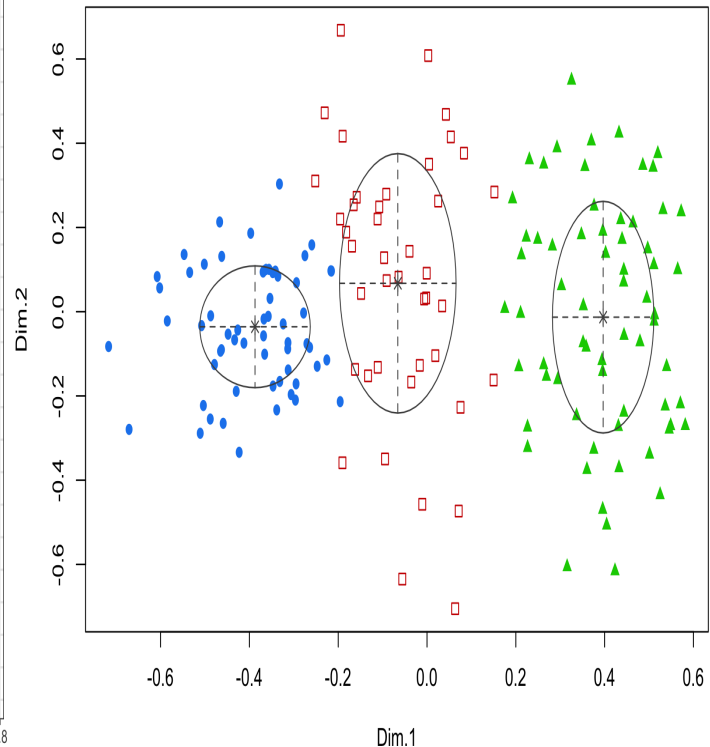
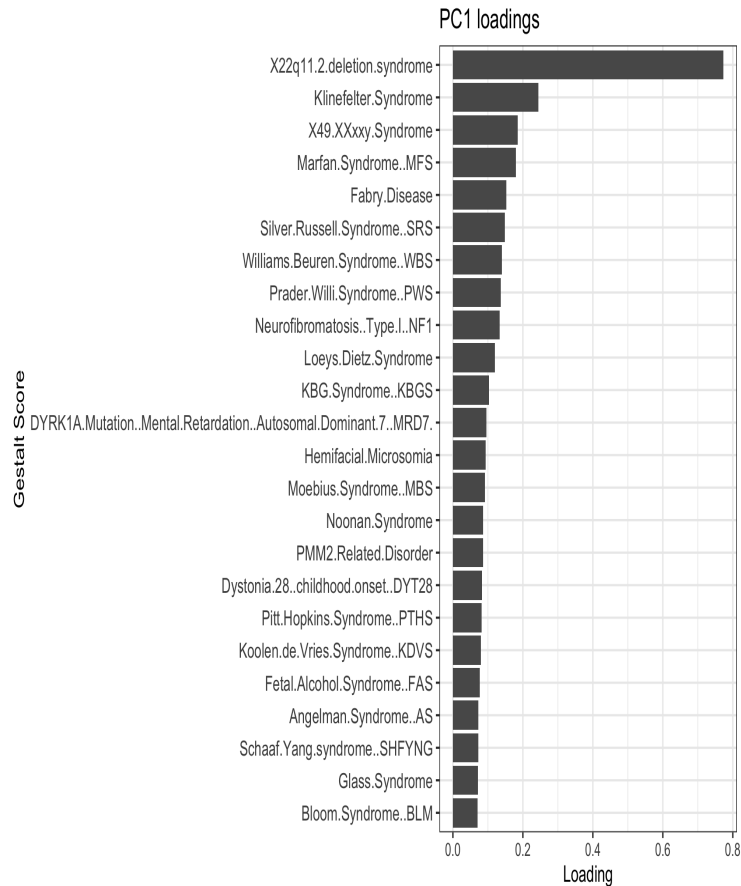
For each analysis section detailed above, an initial LPA was performed on the first  $p$  Principal Components (PC)——  $p = 4, 6, 4$  respectively——where  $p$  was determined by visual examination of the screeplot of eigenvalues (i.e. elbow rule). Clustering results were visually inspected and dimensions of separation were found to be among the first 2 PCs in all cases.

Results shown here are based on the clustering results on the first 2 PCs.

# 22q: Cluster Visualizations

# 22q: Cluster Visualizations

Taking the first 2 PCs, a diagonal, varying volume and shape model with 3 components was highlighted as the best model per the BIC. The clusters are separated by Dimension 1.



## 22q: External Validation

Clustering results were not associated with sex,  $\chi^2(2) = 2.26$ ,  $p = .323$ , or race,  $\chi^2(10) = 5.06$ ,  $p = .887$ .

Clusters did show differences in age, however:

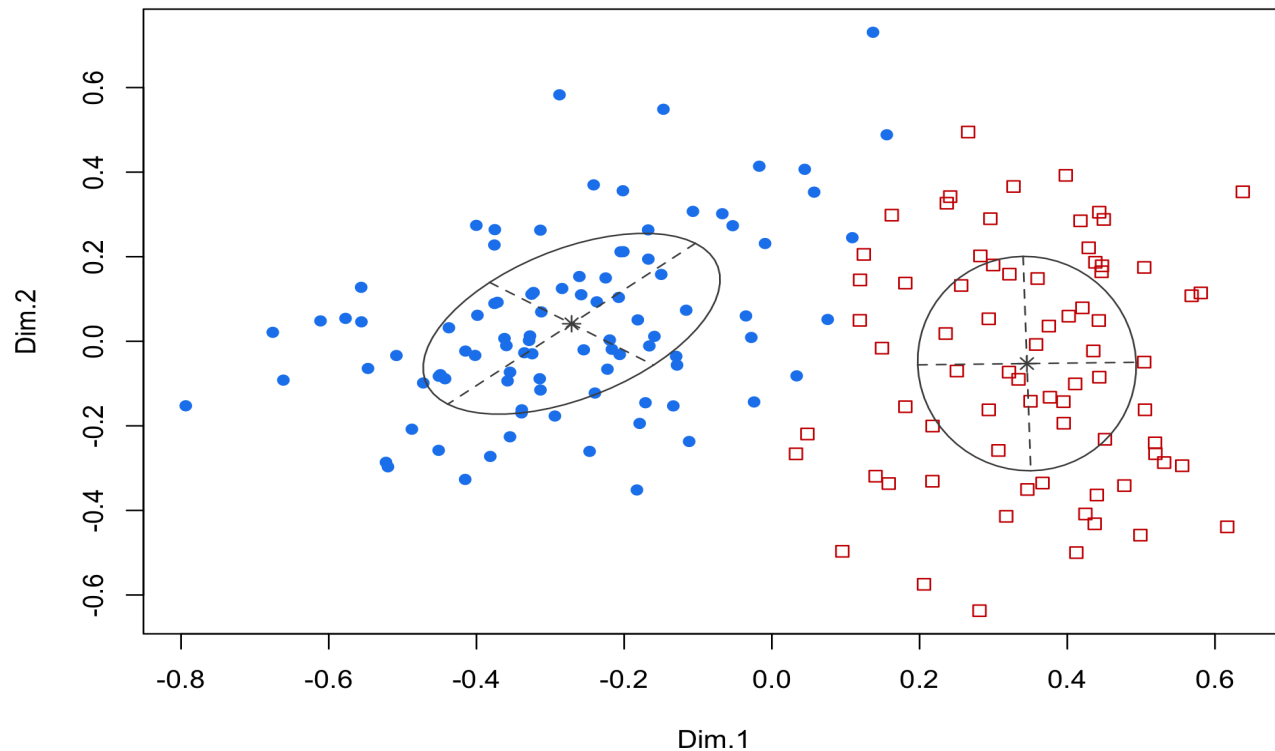
```
##           Effect
## 1 (Intercept) F(1, 165) = 1429.71, p < .001, petasq = .90 ***
## 2 clustering F(2, 165) = 4.71, p = .010, petasq = .05 *
```

And PCs were correlated with age also:

```
##           Dim.1           Dim.2
## age    -0.213218096 -0.2793875100
## p.val   0.005520636  0.0002448886
```

# 22q: Clustering adjusted for Age

When the age effect on the PCs is corrected for, clustering results differed in their number and shape of distribution for in the PCs. Sex and race remain independent from these modified clustering results.



# 22q: Conclusion

- Clustering results showed age effects.
- Dimension 1 best separated the clusters in the age-adjusted and non-adjusted results.

For follow up:

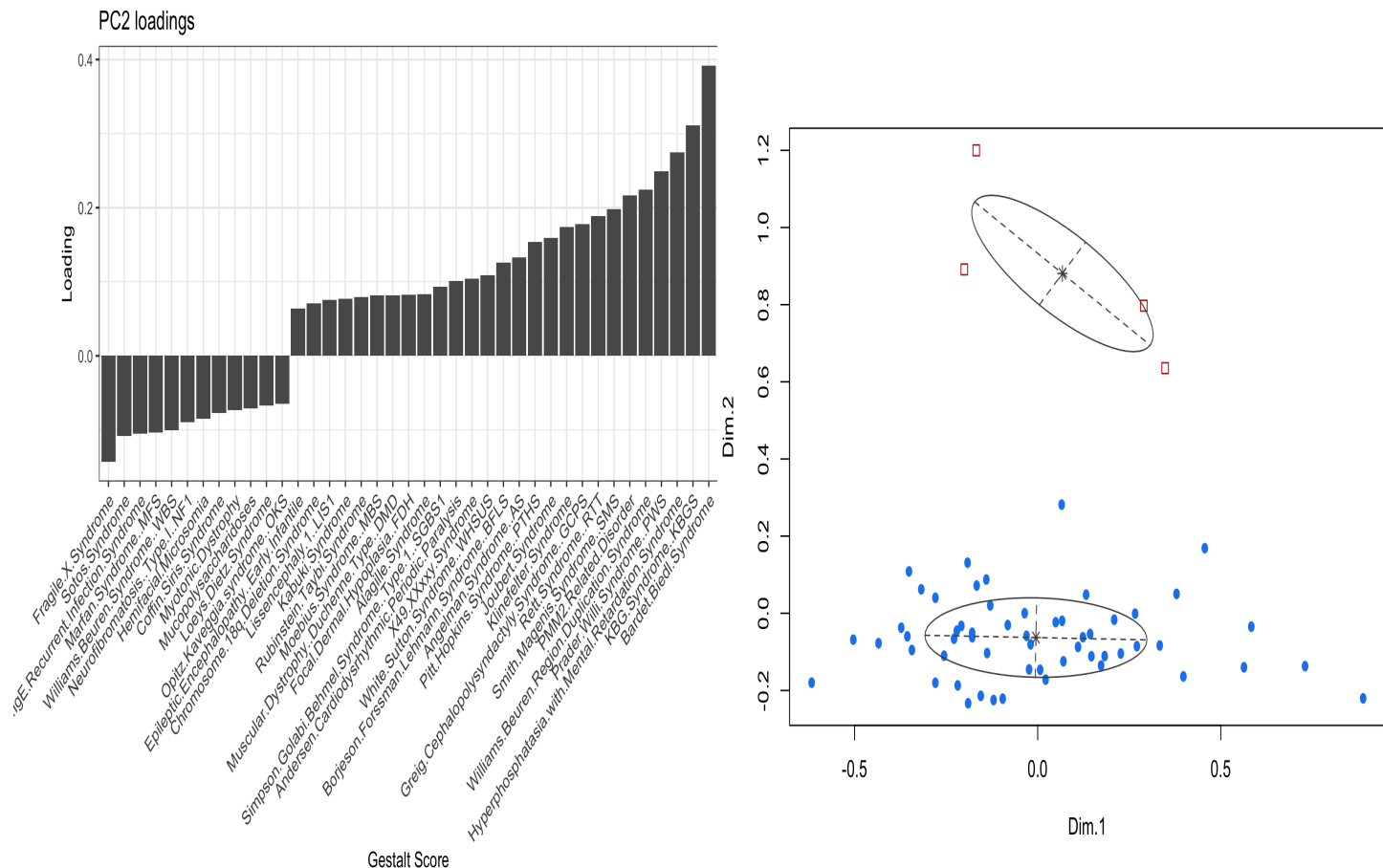
- Examine cluster differences in IQ, psychotic features and facial measurements.

# PS: Cluster visualization



## PS: Cluster visualization

A ellipsoidal, equal volume and shape model with 2 components was highlighted as the best model by the LPA. The clusters are separated by Dimension 2.



# PS: External Validation

Clustering results were not associated with sex,  $\chi^2(1) = 0.64$ ,  $p = .423$  or race  $\chi^2(1) = 0.64$ ,  $p = .423$ .

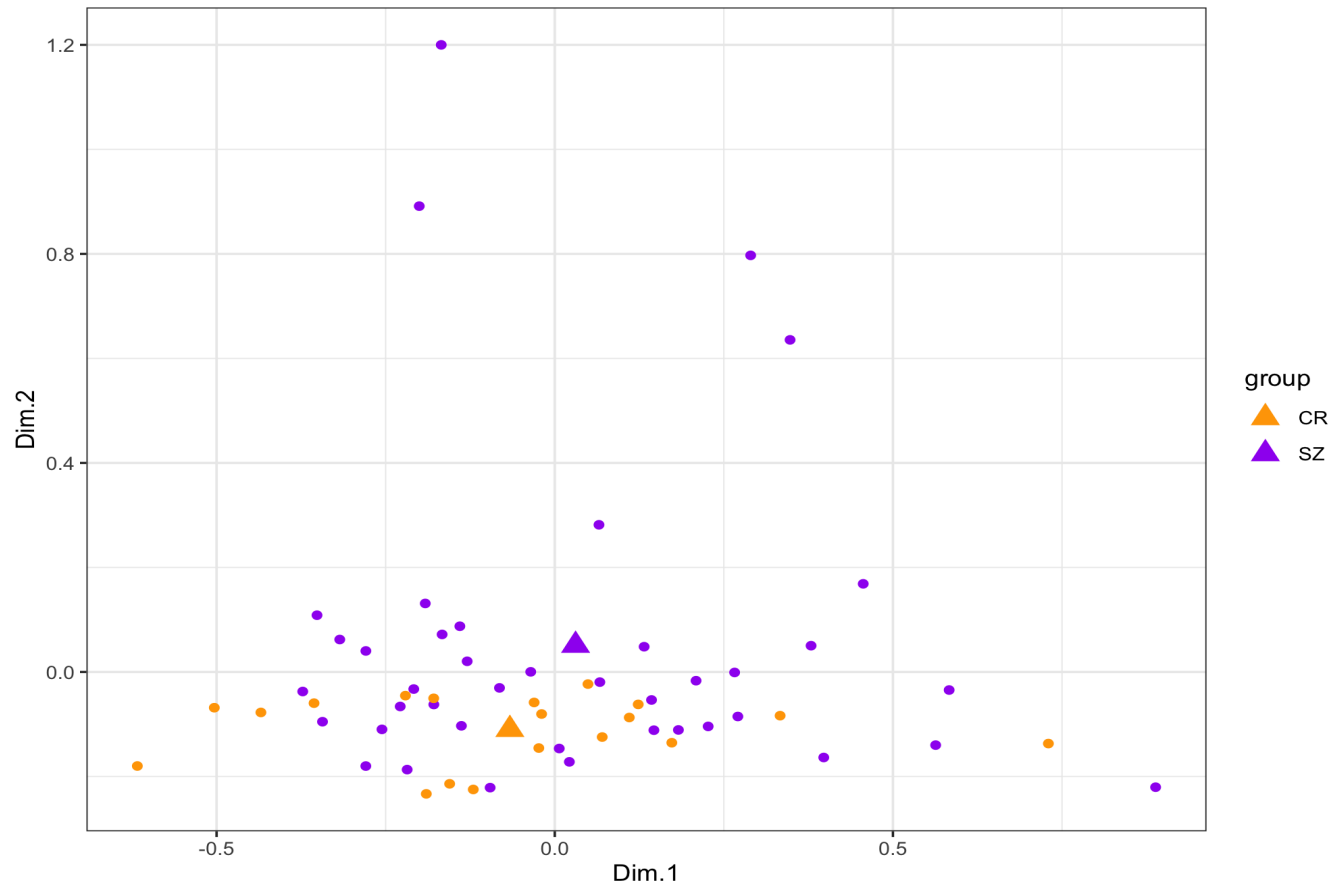
Clusters did not show differences in age :

```
##          Effect
## 1 (Intercept) F(1, 58) = 598.94, p < .001, petasq = .91 ***
## 2 clustering F(1, 58) = 0.09, p = .760, petasq < .01
```

Psychosis Spectrum sub-groups Schizophrenia (SZ) and Clinical Risk (CR) were not correlated with clustering results  $\chi^2(1) = 0.73$ ,  $p = .394$

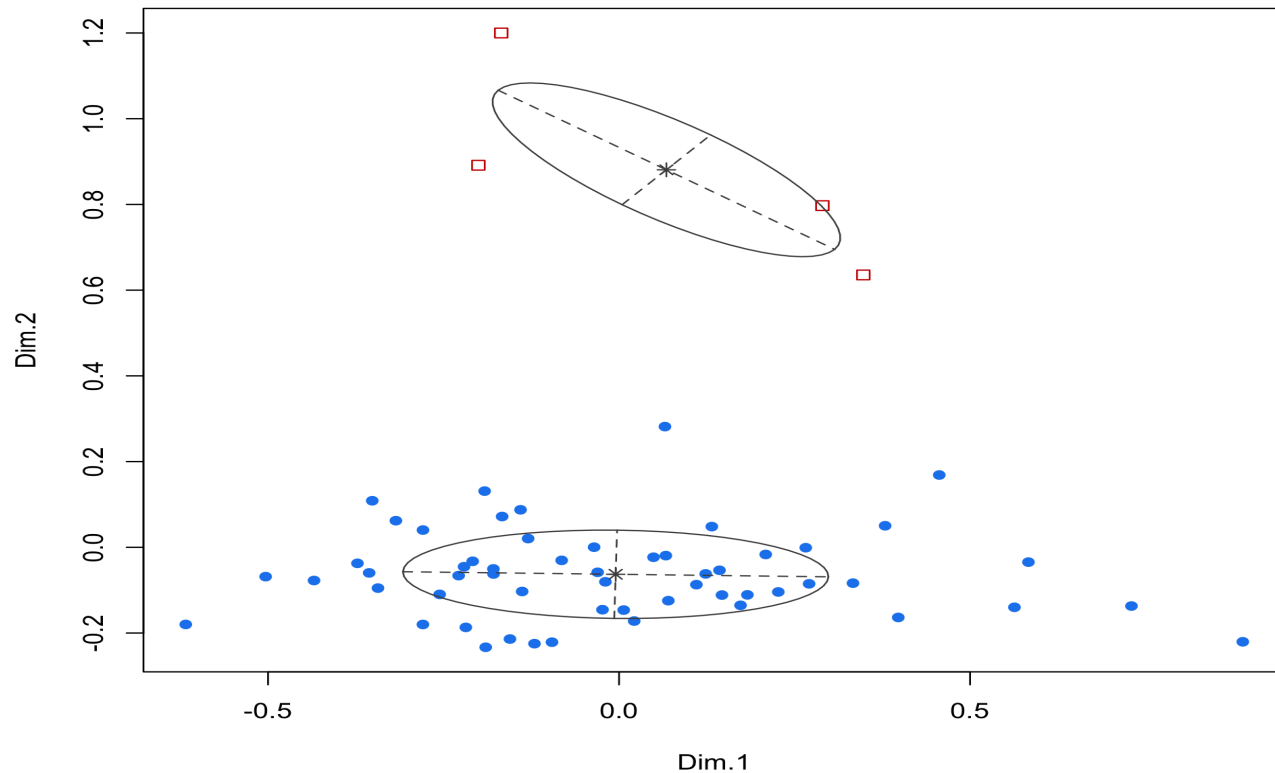
# PS: PC differences

There were significant differences between the SZ and CR groups on PC 2,  $t(58) = 2.28$ ,  $p = .026$ ,  $d = 0.59$ , but not on PC 1,  $t(58) = 1.16$ ,  $p = .251$ ,  $d = 0.30$ .



# PS: Clustering adjusted for Age

Principal component 1 was associated with age  $r = 0.36$ ,  $p = 0.01$ . However, adjusting PC scores for age did not alter the clustering results.



# PS: Conclusions

- Clustering results were sensitive to outliers.
- Dimension 2 is more informative than clustering results.

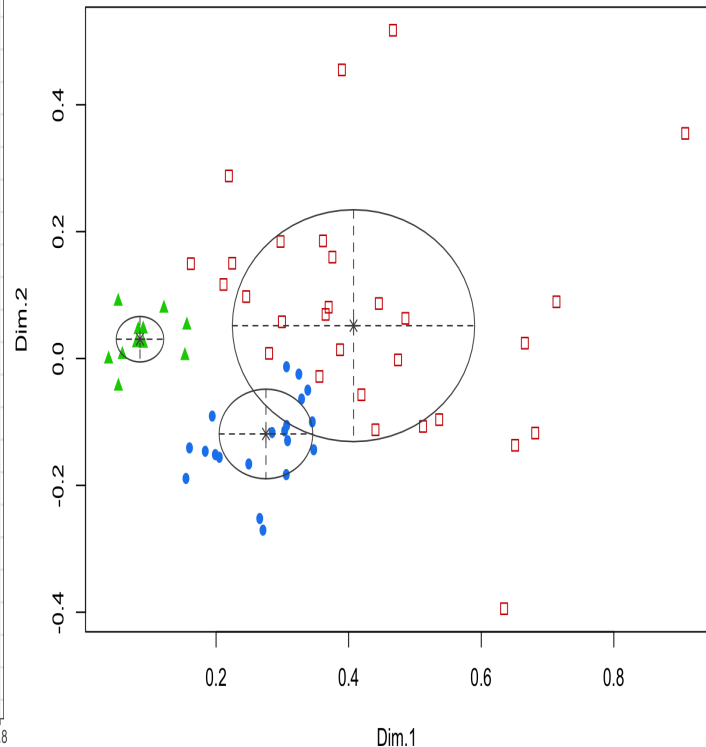
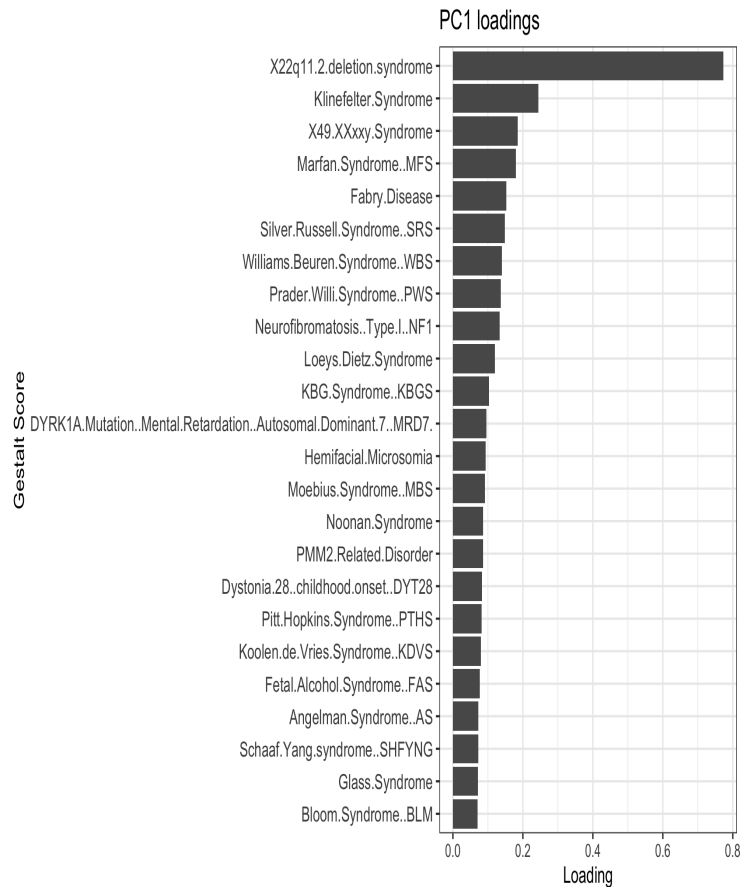
For follow up:

- Examine cluster differences with IQ, psychotic features and facial measurements.

# projPS: Cluster Visualization

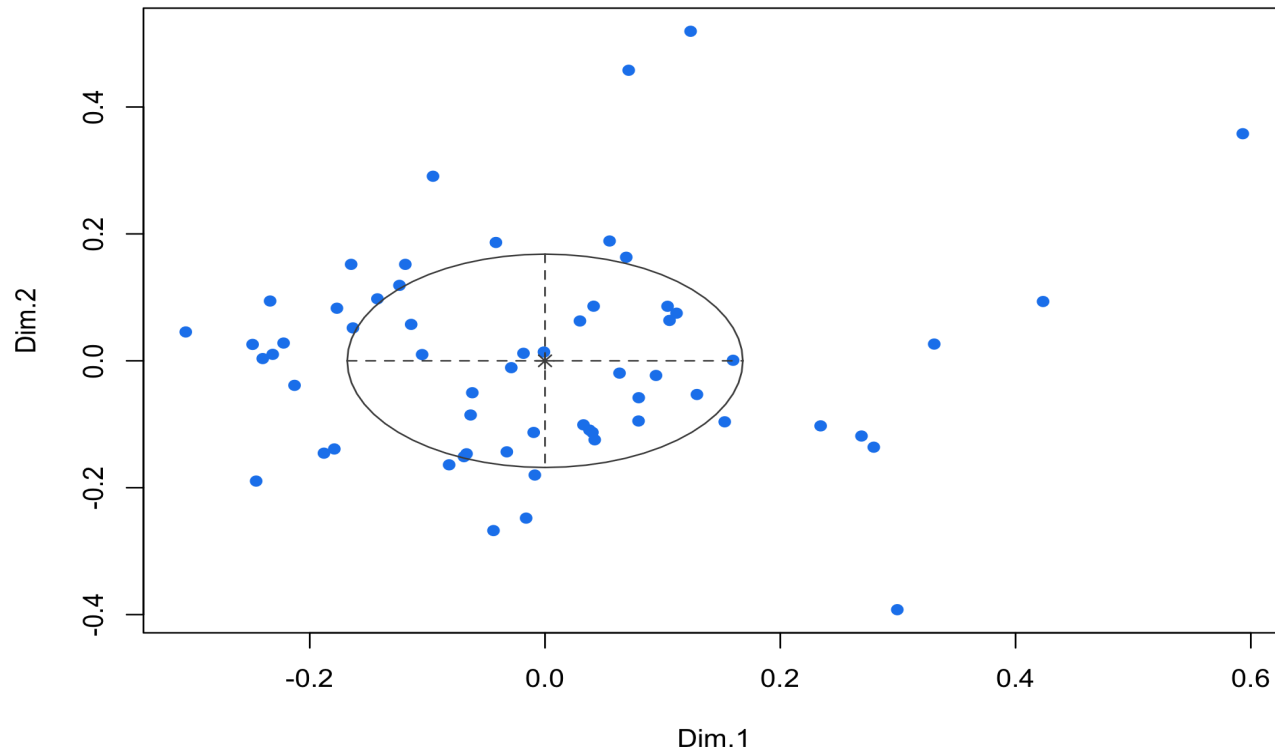
# projPS: Cluster Visualization

A spherical, varying volume model with 3 components was highlighted as the best model. The clusters are separated by Dimension 1.



# projPS: Clustering adjusted for Age

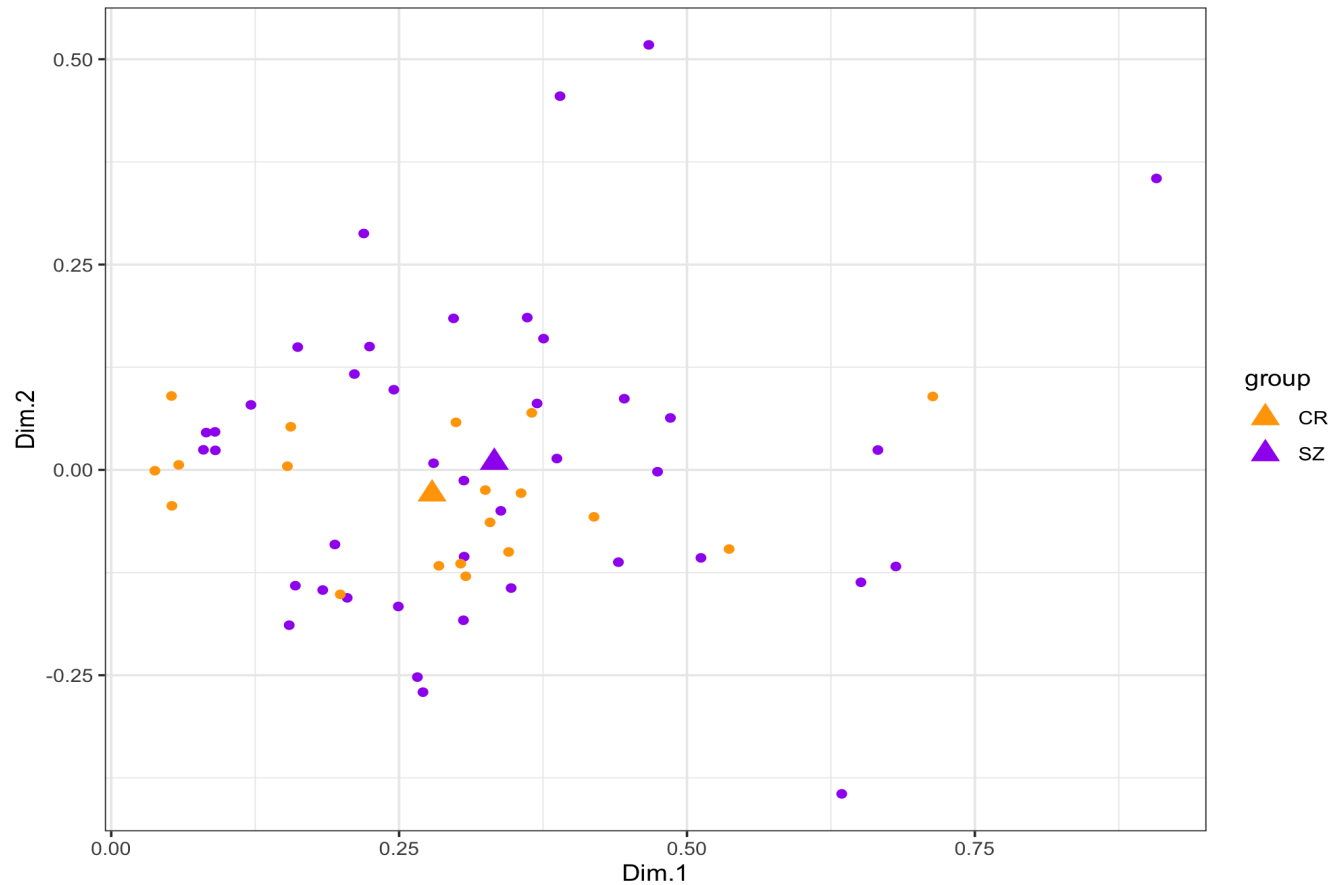
When the age effect is corrected for, the clustering solution collapses into a singular distribution suggesting spurious effects associated with age in the original results.





# projPS: PC differences

There were no significant differences between SZ and CR groups on PC1,  $t(58) = 1.07$ ,  $p = .291$ ,  $d = 0.28$ , or PC2,  $t(58) = 0.85$ ,  $p = .397$ ,  $d = 0.22$ .



# projPS: Conclusions

- Clustering results were predicated primarily on age effects suggesting spurious results.

# Follow up

The analysis of PS projected scores reveals there can be strong age effects at play in the 22q feature space. Therefore, validating 22q11.2 deletion clusters with the remaining variables——psychotic features, IQ, and facial measurements——would be appropriate to aid interpretation.

In the case of PS clusters, examining psychotic symptoms and their interactions could be informative for validating any potential cluster structure.

# References

McLachlan, G. J., Lee, S. X., & Rathnayake, S. I. (2019). Finite mixture models. *Annual Review of Statistics and Its Application*, 6, 355–378. <https://doi.org/10.1146/annurev-statistics-031017-100325>

Scrucca L., Fop M., Murphy T. B., Raftery A.E. (2016). “mclust 5: clustering, classification and density estimation using Gaussian finite mixture models.” *The R Journal*, 8(1), 289–317. <https://doi.org/10.32614/RJ-2016-021>.