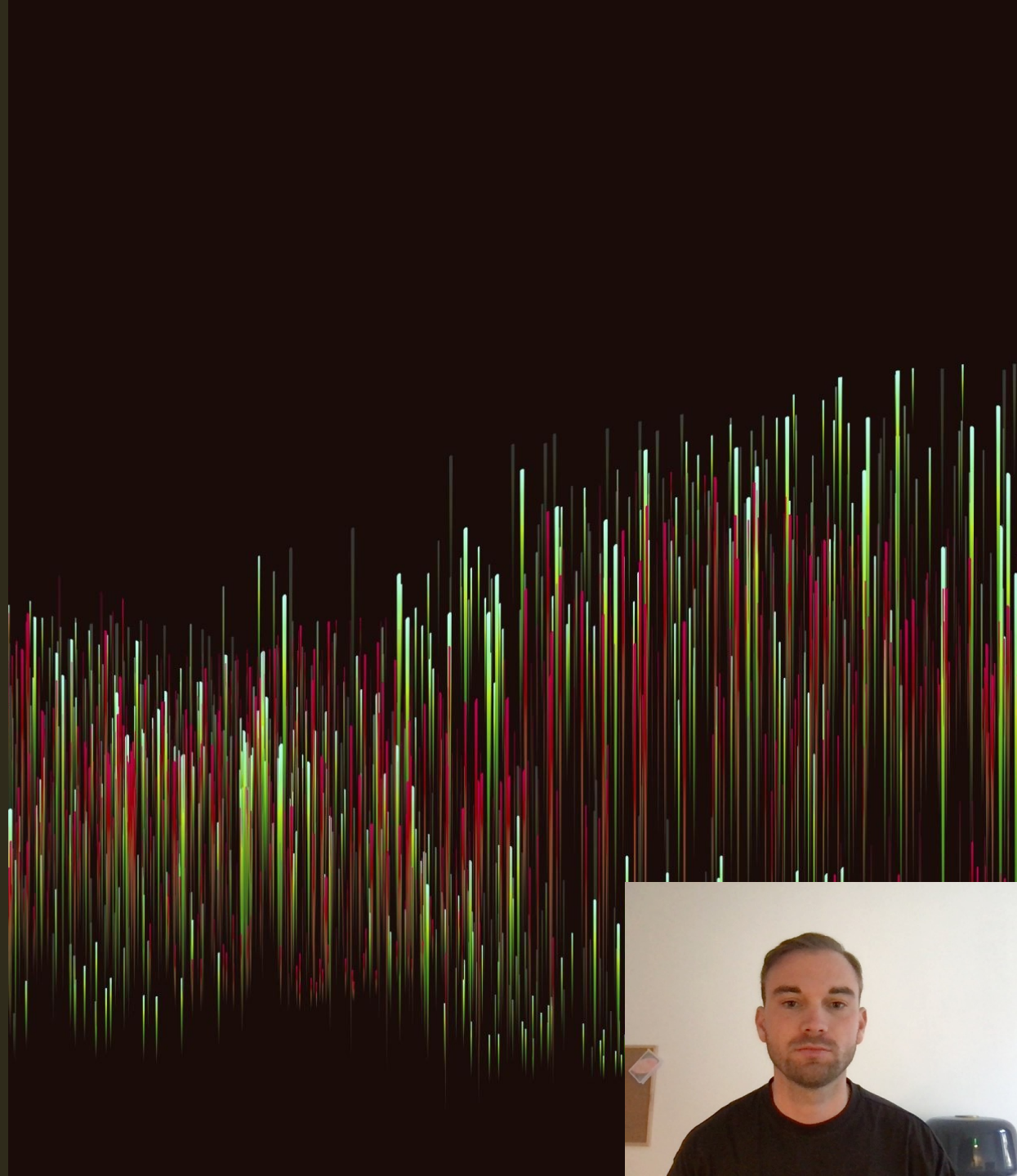
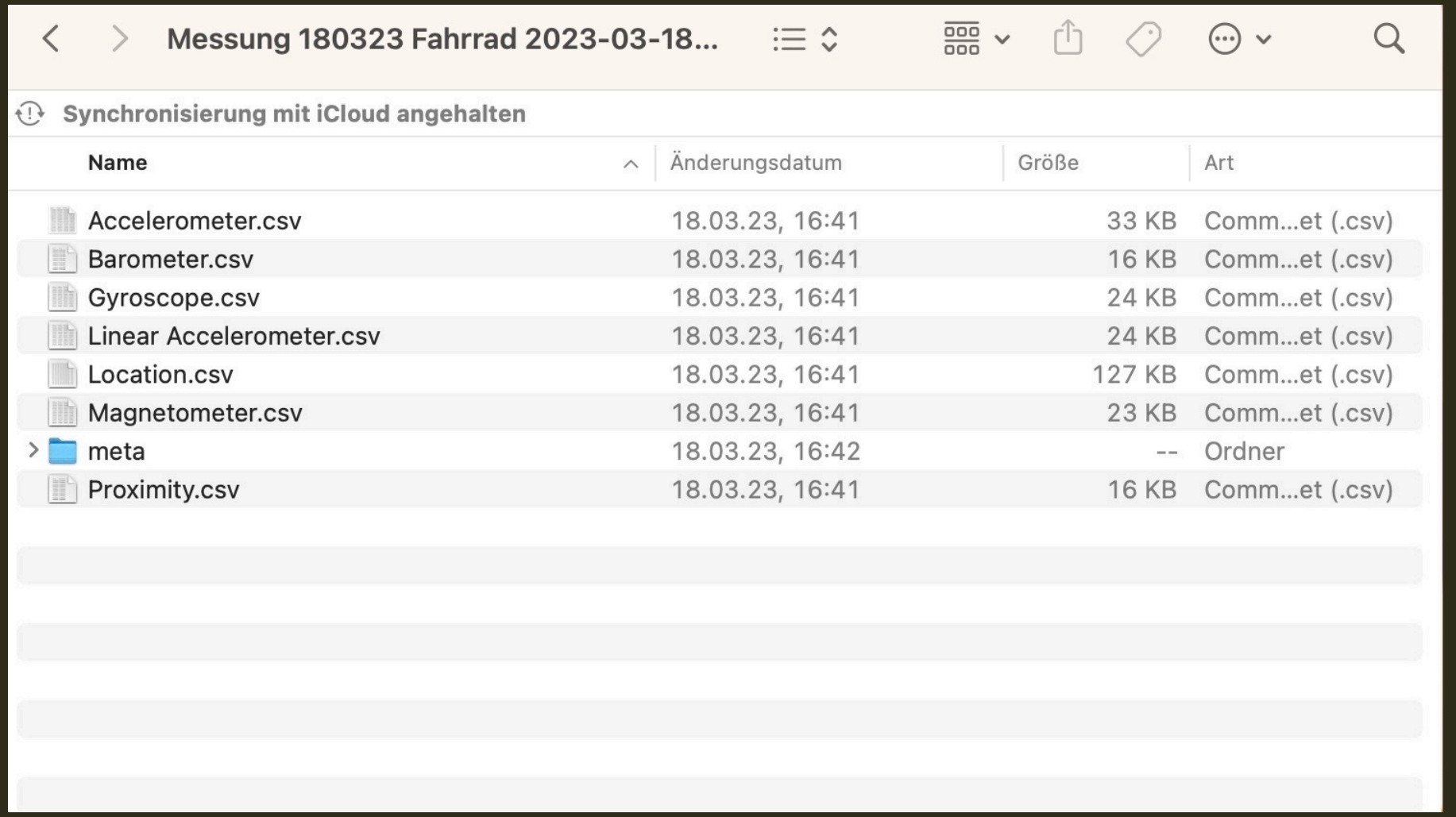
Applications of Datascience



Preprocessing

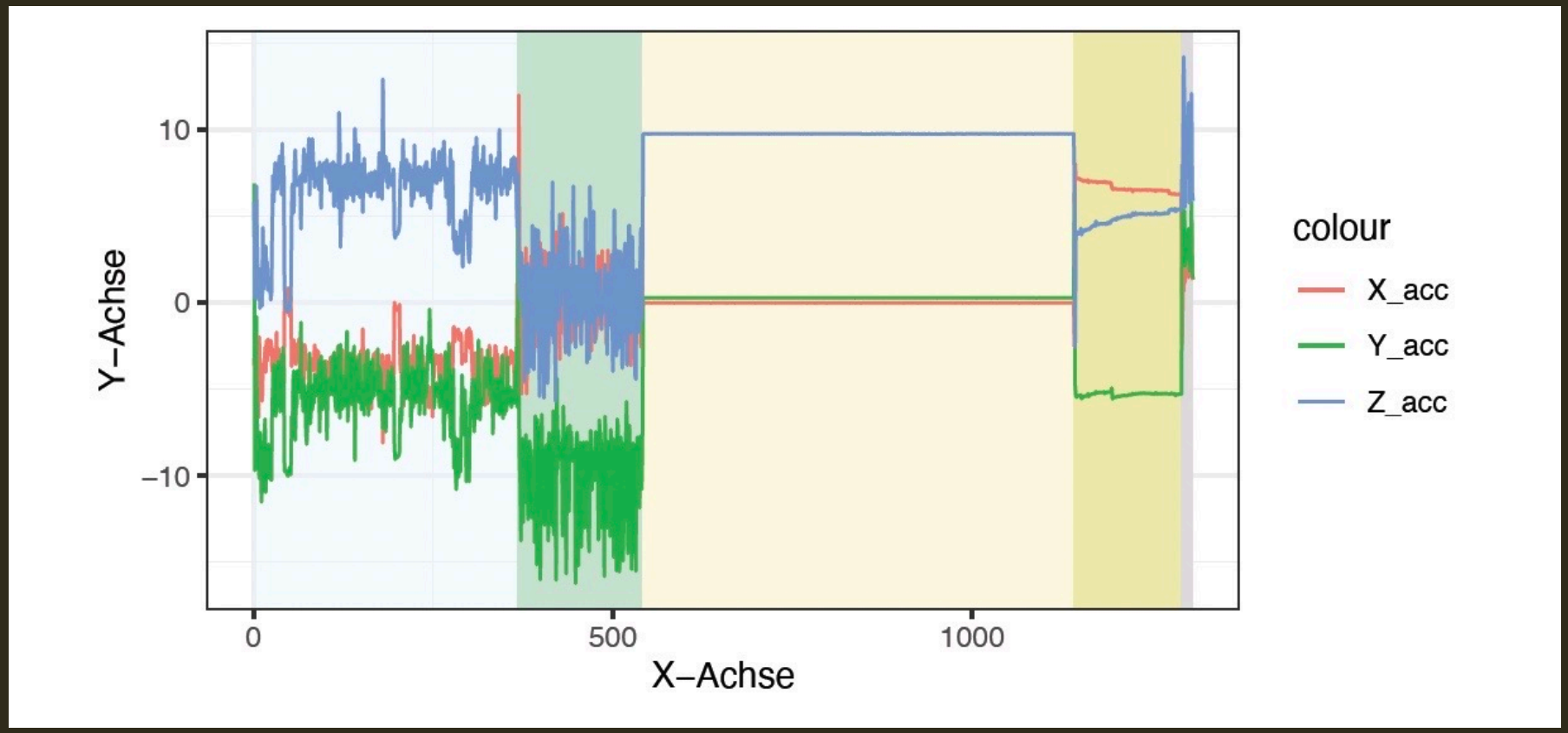


Dataframe I worked with

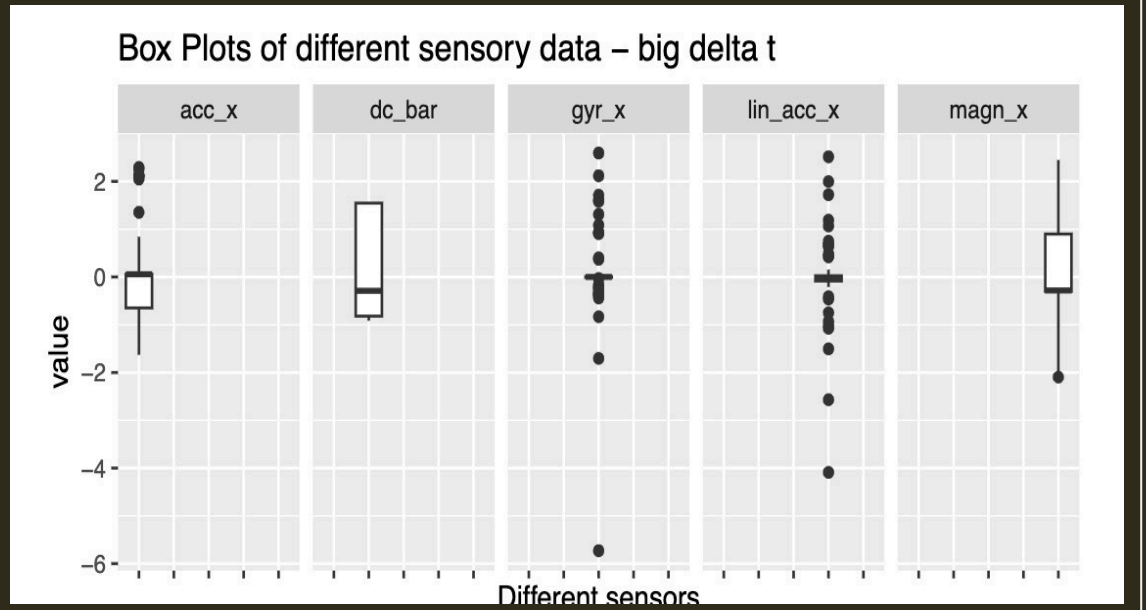
Head of the Dataframe with 1 column for every sensor

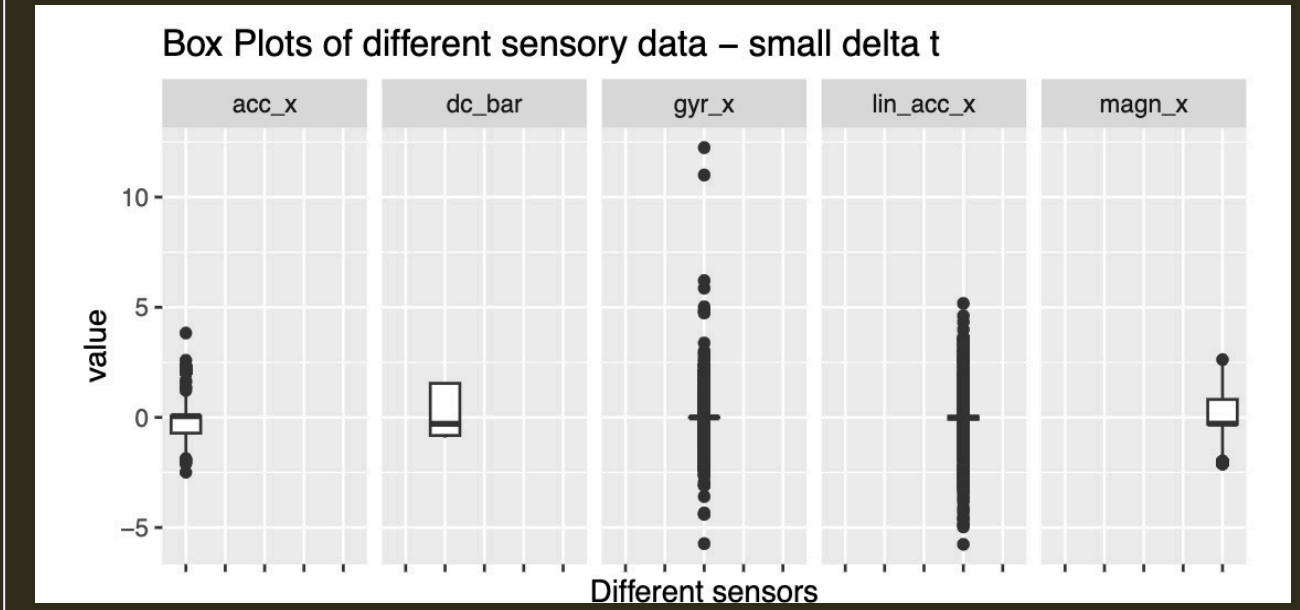
X_acc	X_hPa	X_rad	X_linacc	X_µT	Activity
-3.656746	993.4352	5.40708780	-3.9807026	-28.163292	fahrrad
-1.297652	993.4512	0.19785871	0.4741257	6.101456	fahrrad
-2.451302	993.4512	-0.09114313	-0.1024862	4.411934	fahrrad
-6.856941	993.4512	-0.69596463	2.1904757	3.494606	fahrrad
-2.876269	993.4512	0.22384980	-0.7286204	-4.406799	fahrrad

Plot of acceleration data

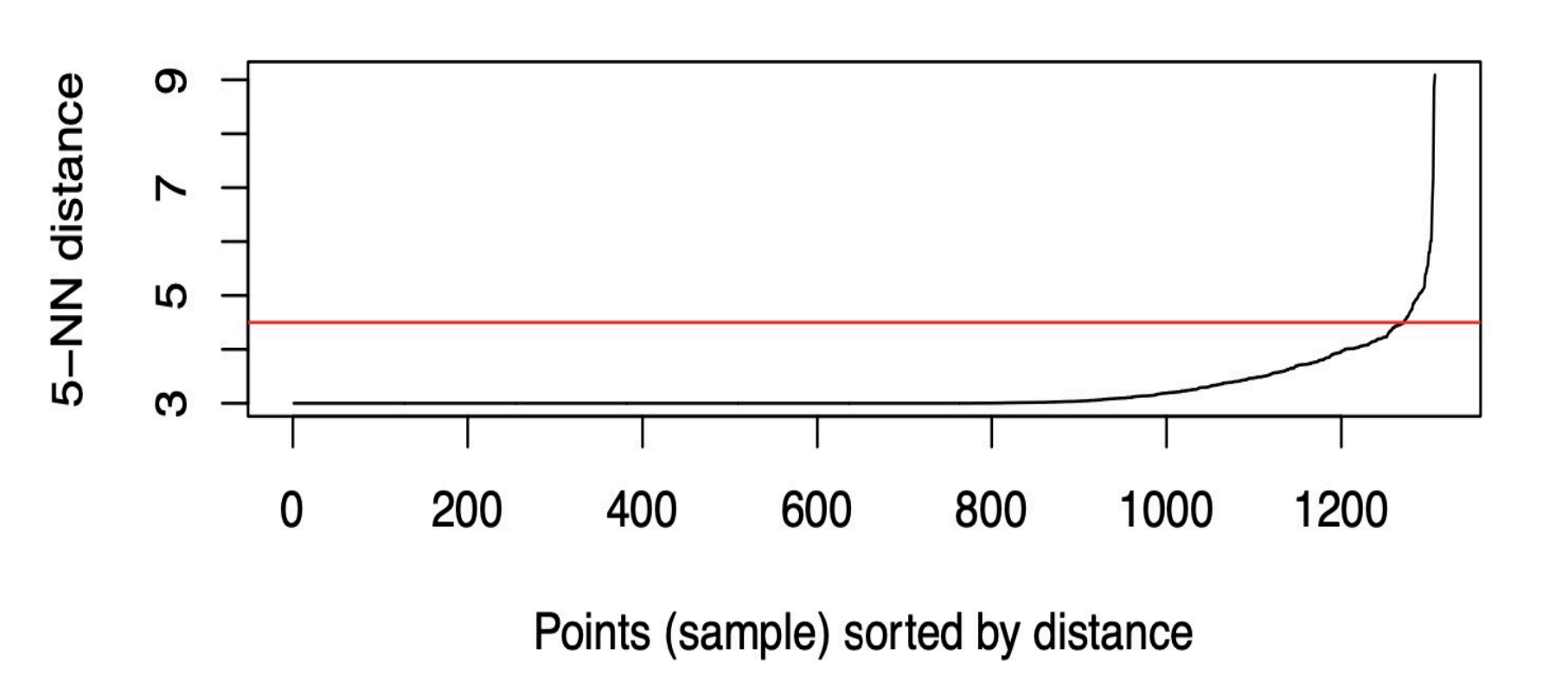


Difference in delta t

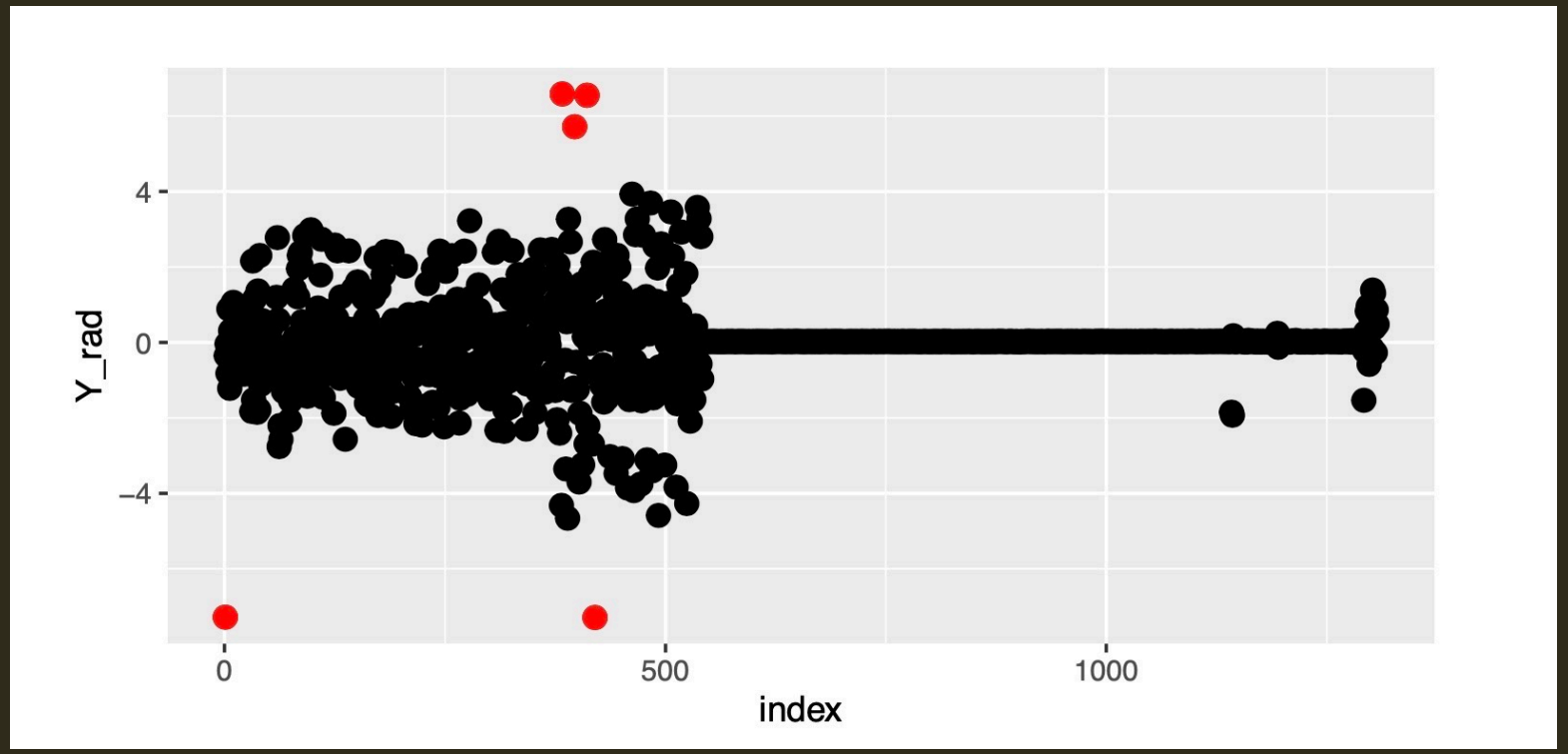




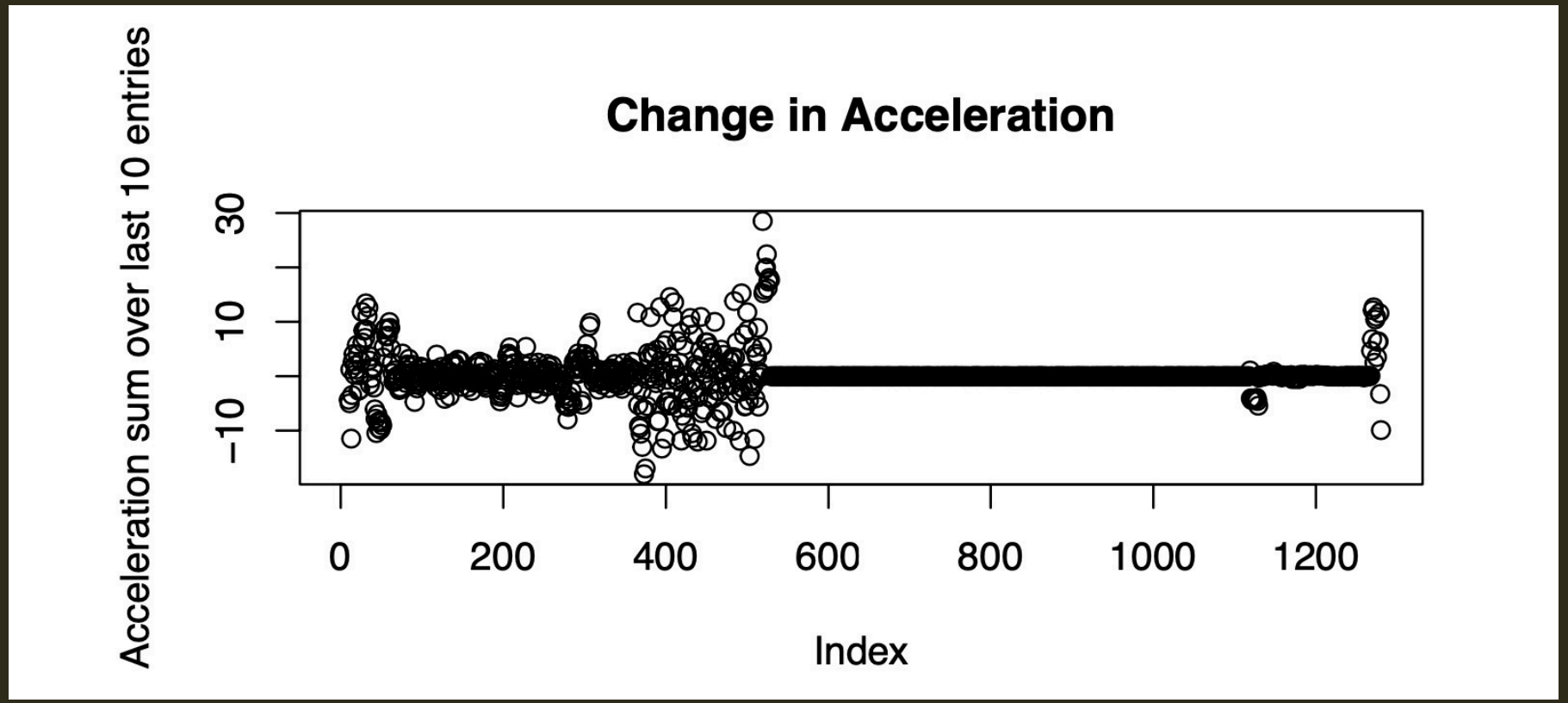
Outlier detection with dbscan - 1



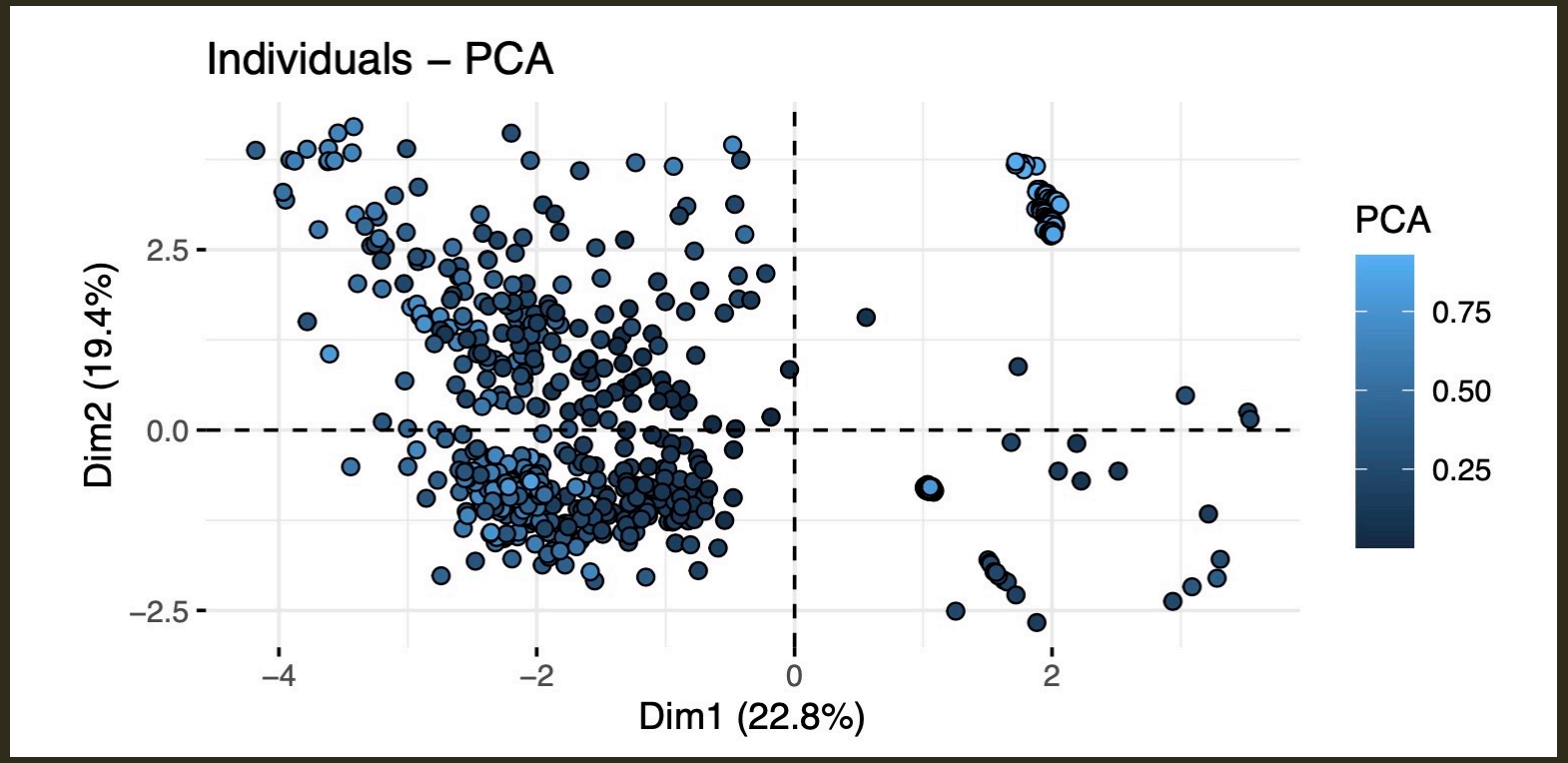
Outlier detection with dbscan - 2



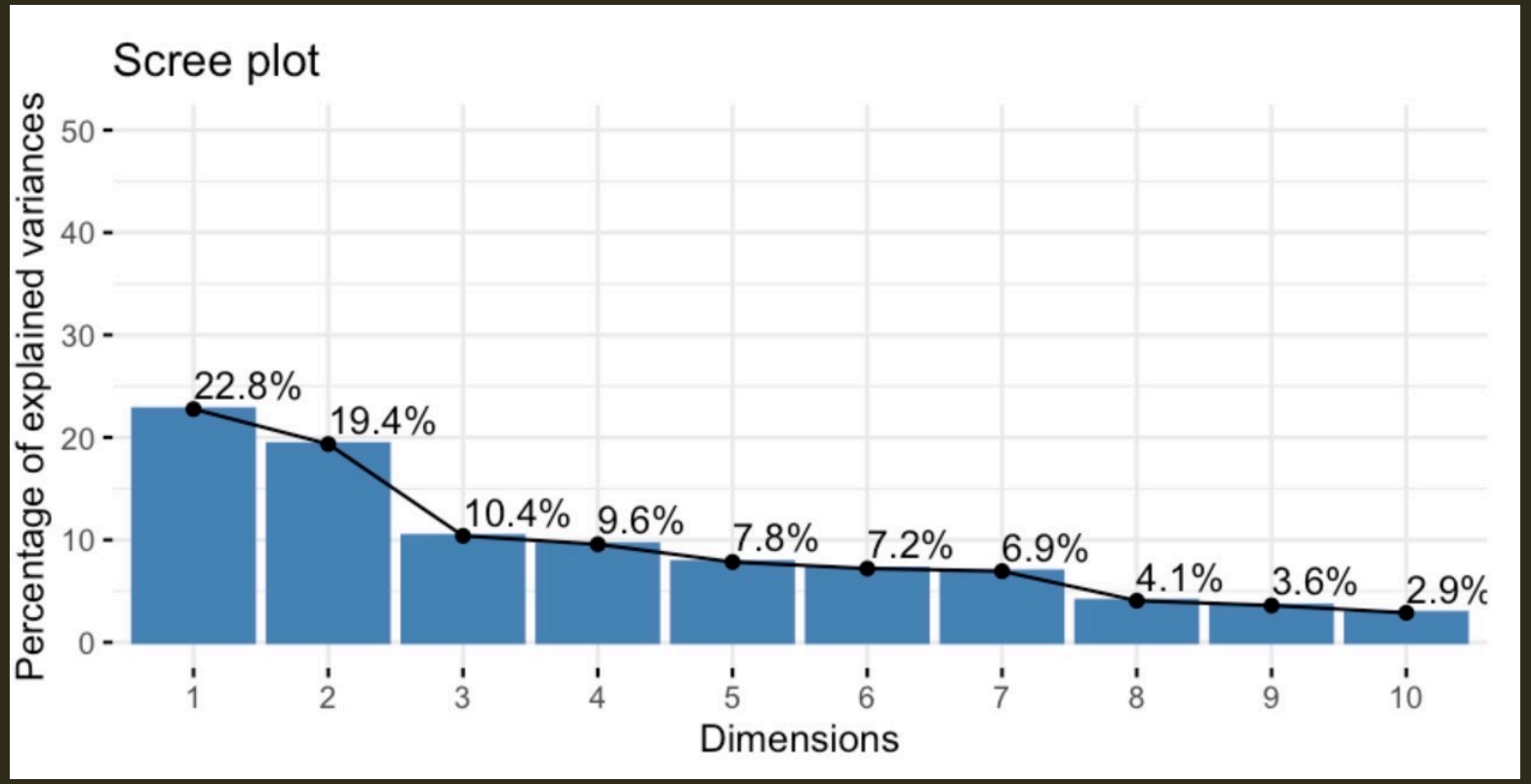
Feature engineering – Temporal Aggregation of acceleration data



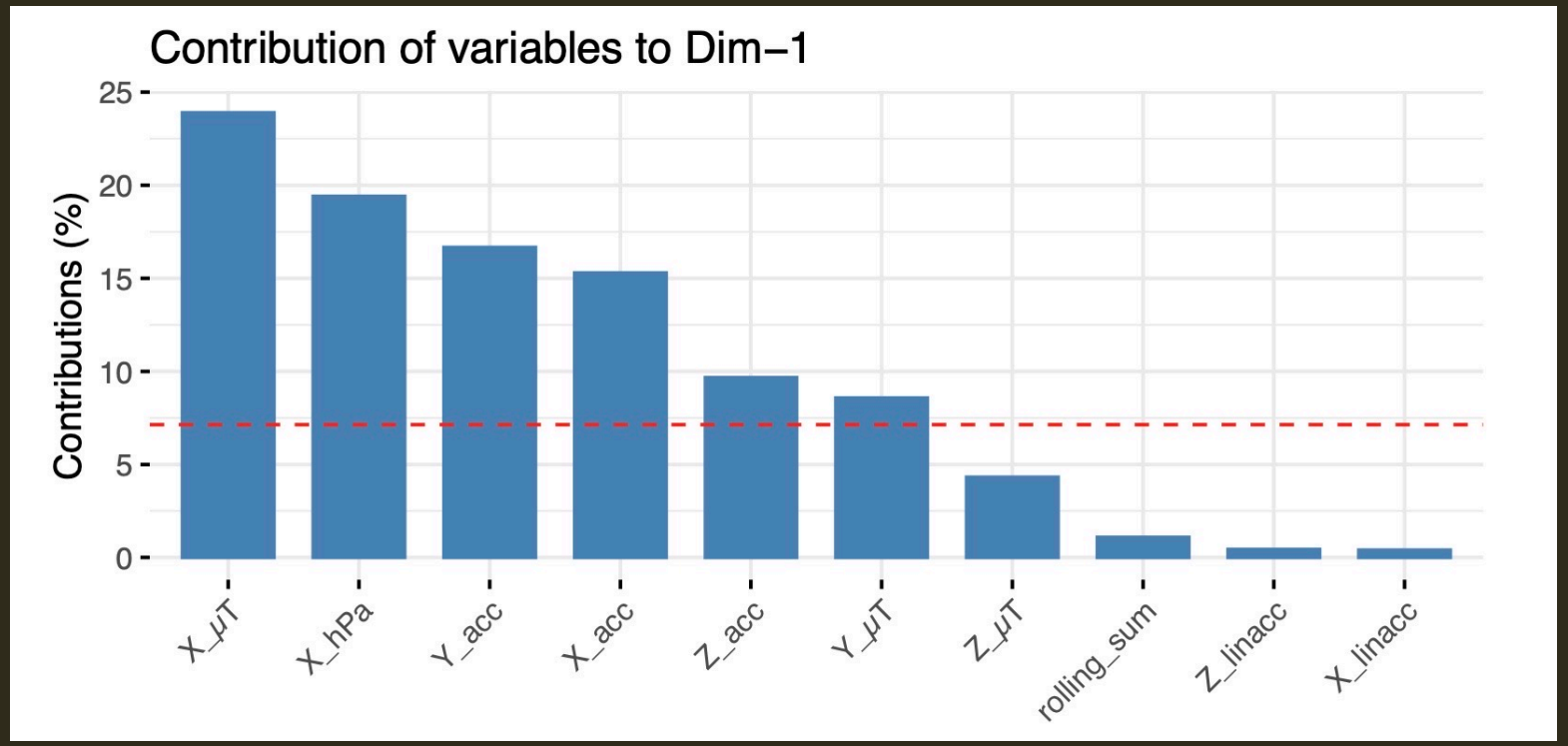
PCA to get a better understanding of the data



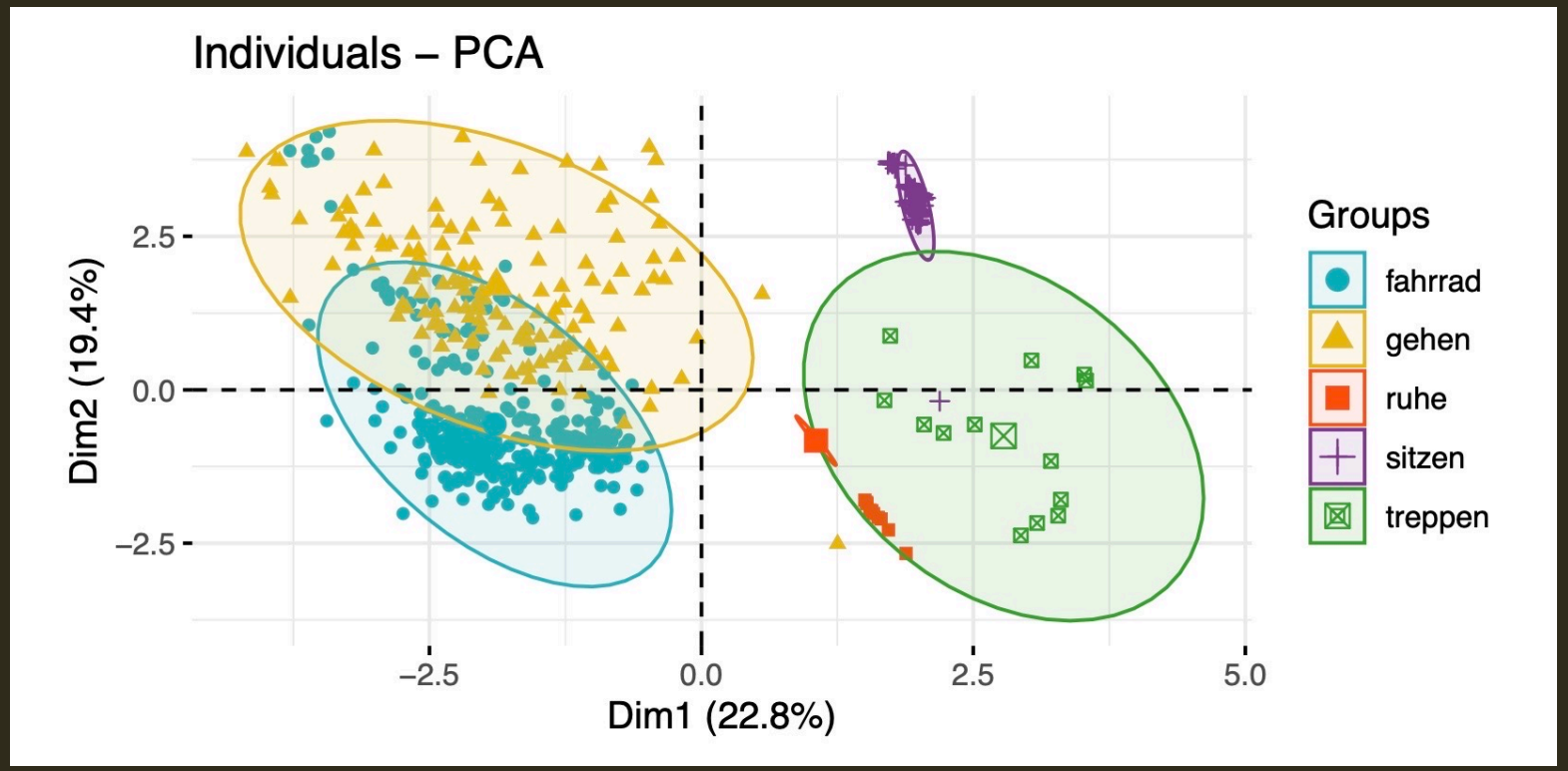
Scree plot


Scree plot
Percentage of explained variances
50
40
30
20
10
0
22.8%
19.4%
10.4%
9.6%
7.8%
7.2%
6.9%
4.1%
3.6%
2.9%
1
2
3
4
5
6
7
8
9
10
Dimensions

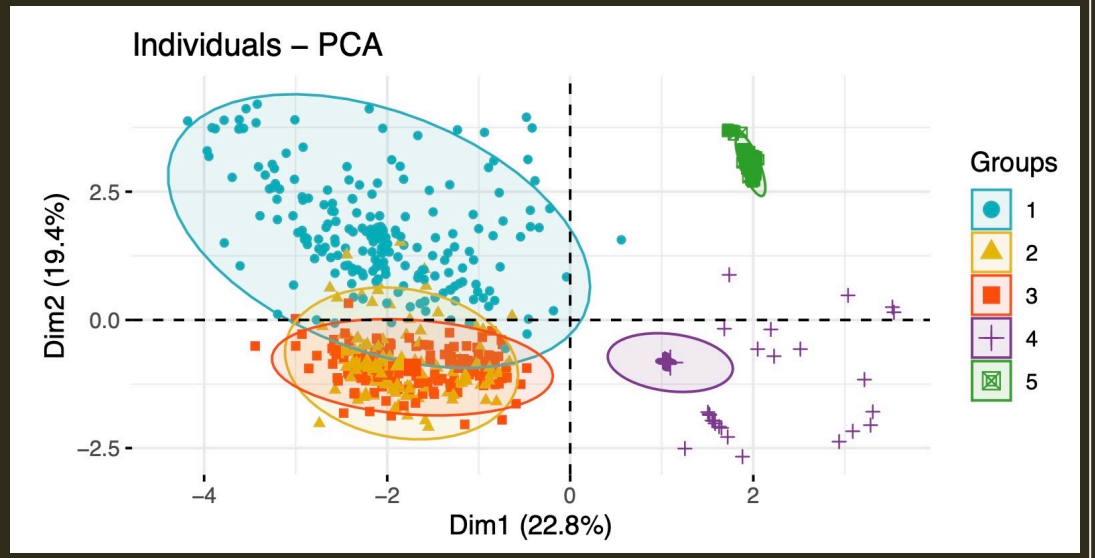
Contribution to principal component

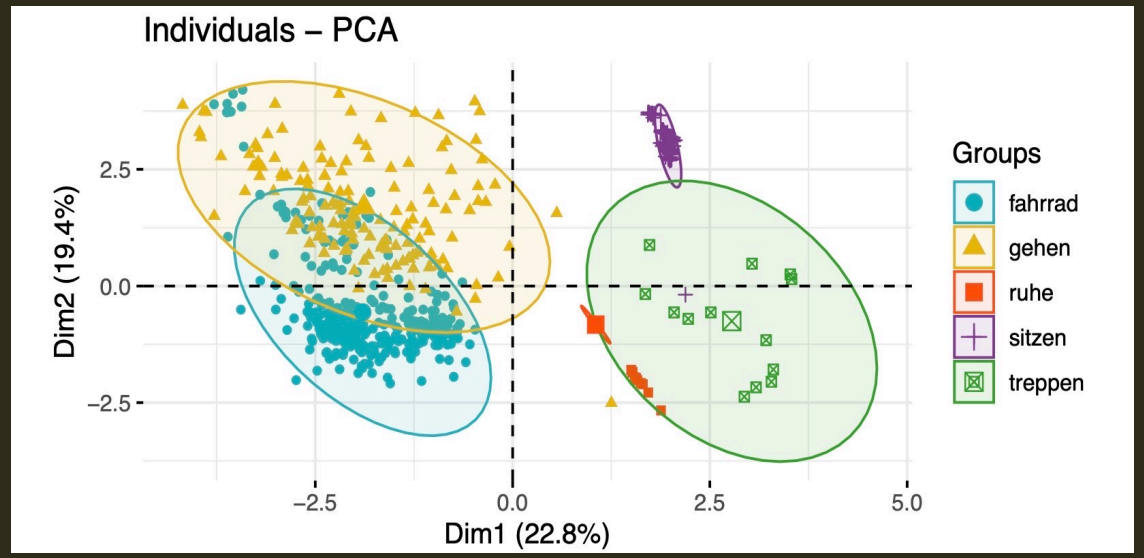


Real Cluster

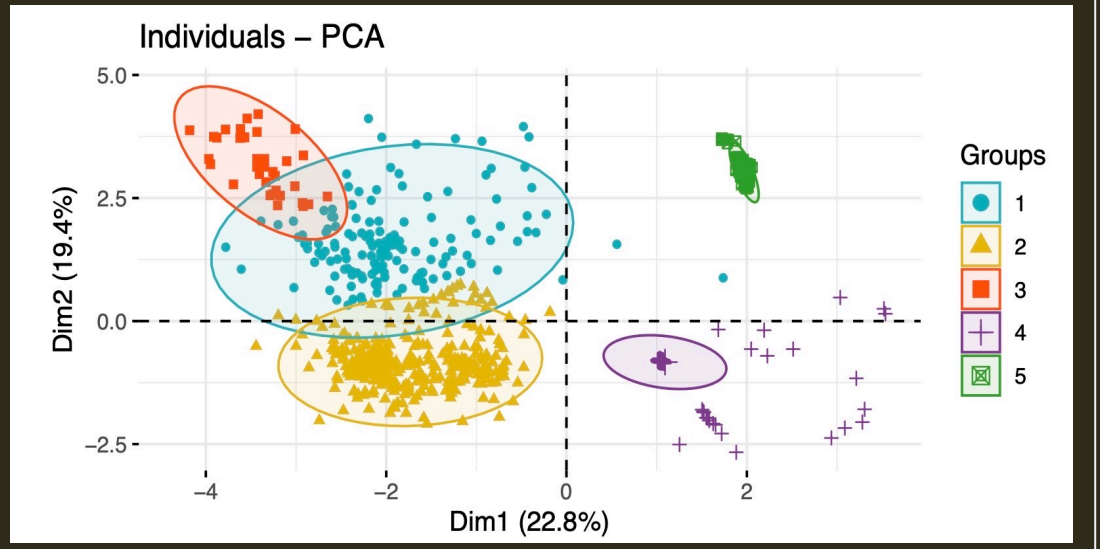


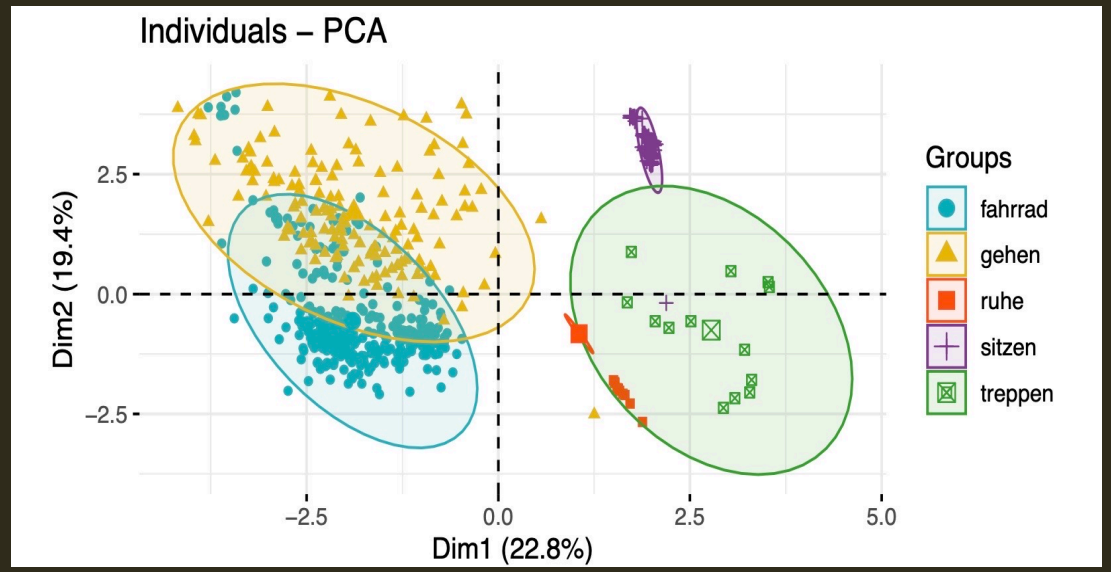
Hierarchial clustering - 1



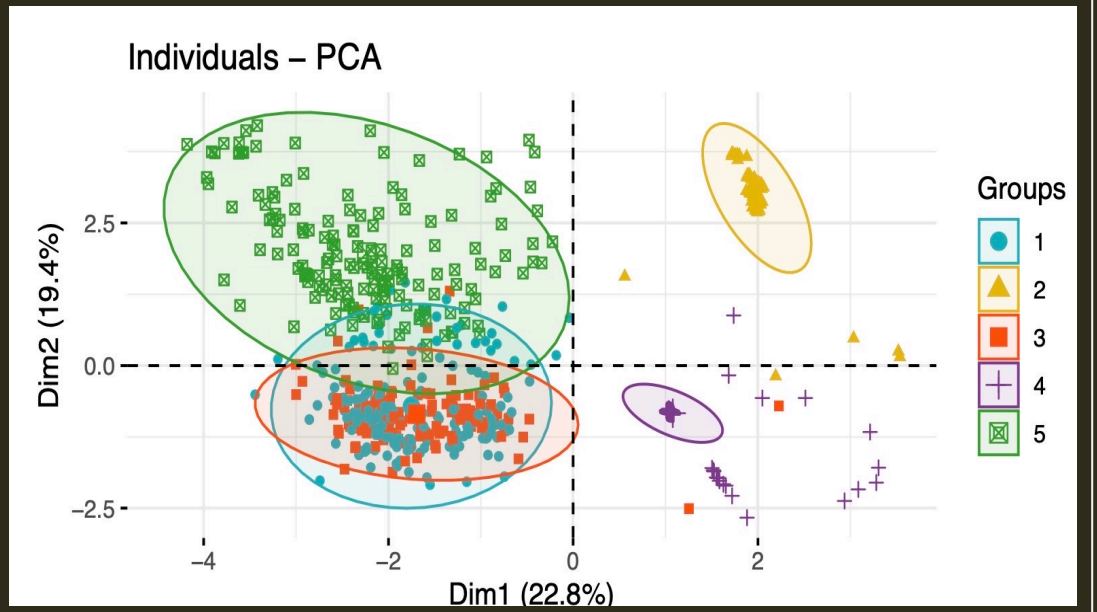


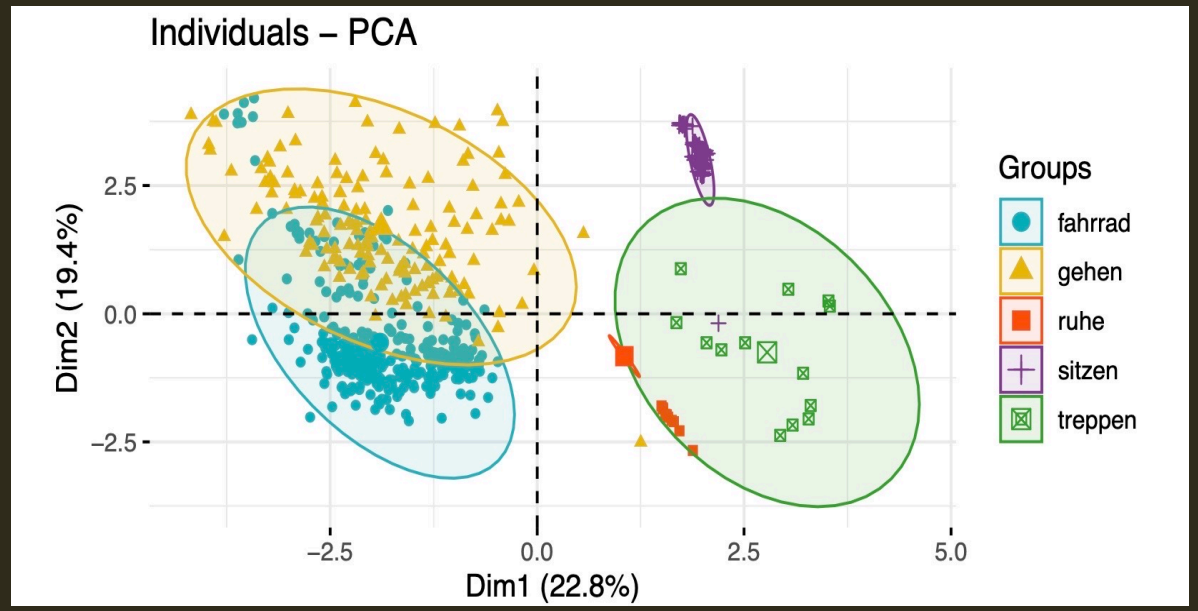
Hierarchial clustering - 2



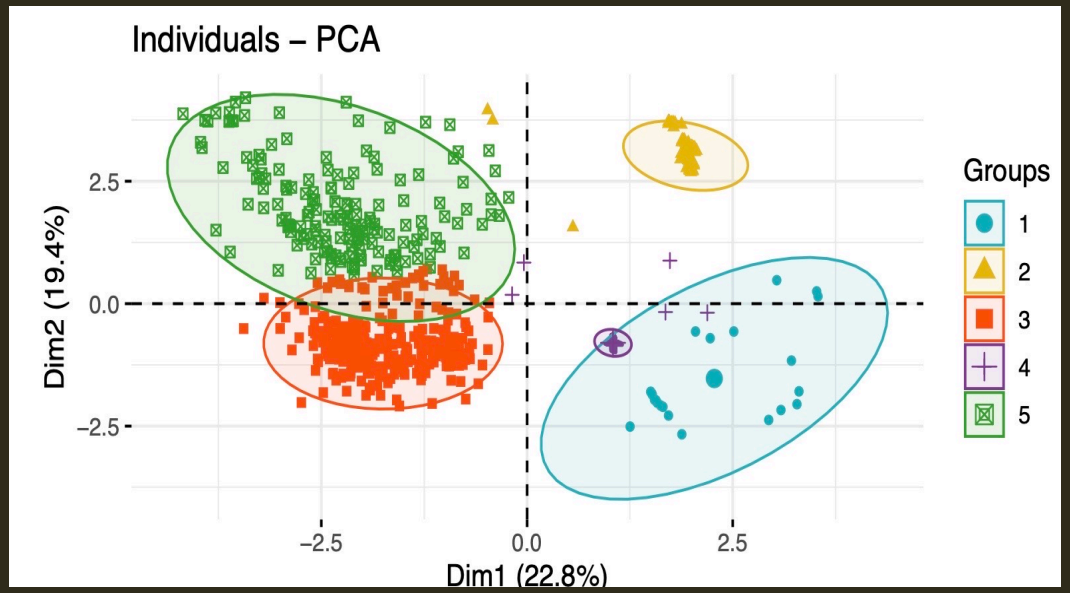


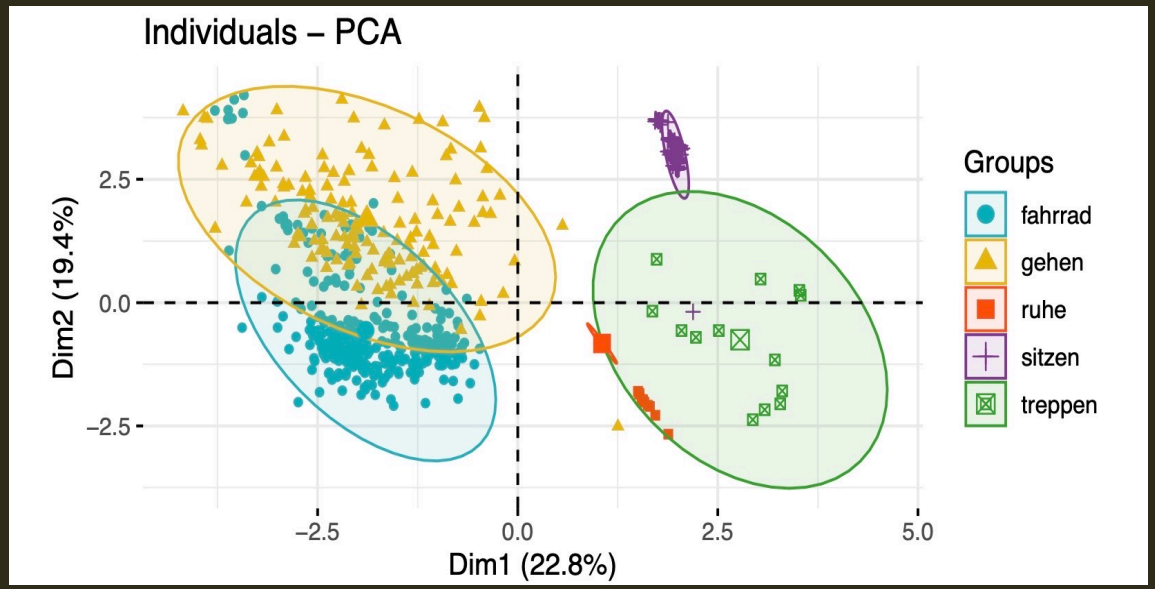
K-means clustering - 1



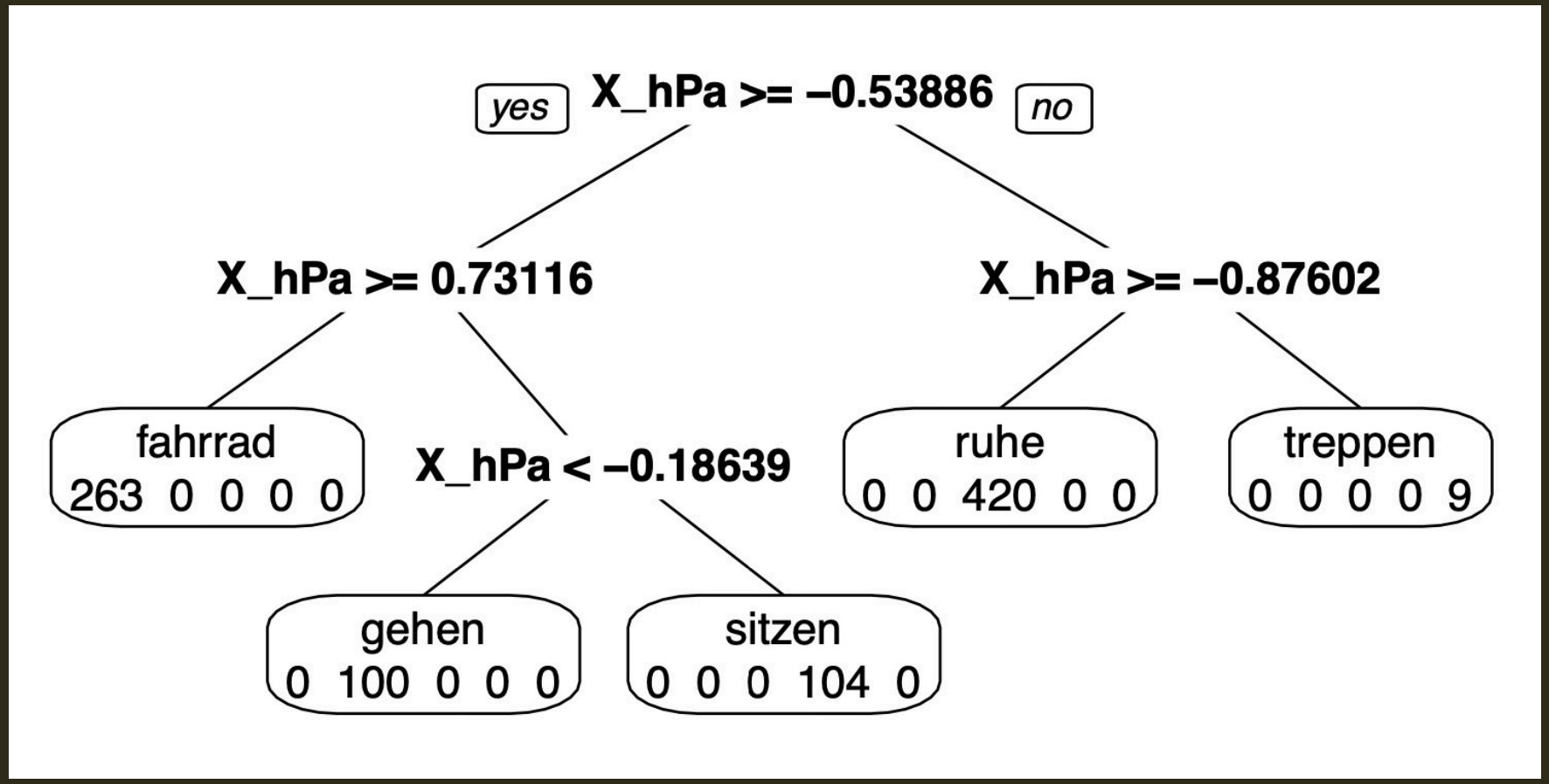


K-means clustering - 2

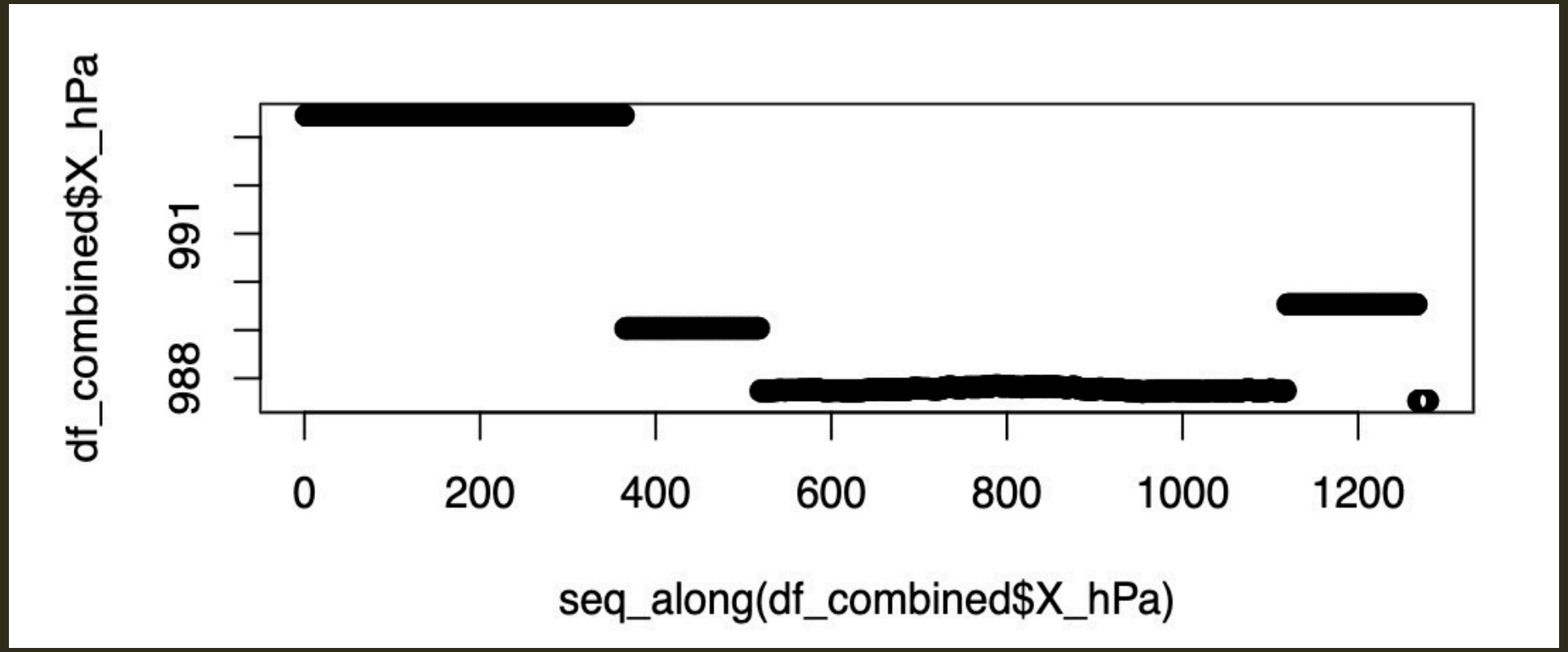


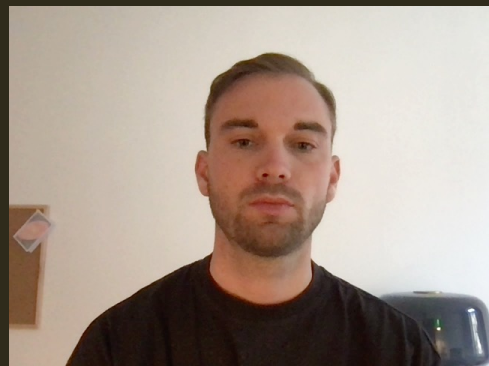


Classification tree – original data



Exploring reasons





Classification tree – confusion matrix

	fahrrad	gehen	ruhe	sitzen	treppen
fahrrad	101	0	0	0	0
gehen	0	54	0	0	0
ruhe	0	0	180	0	0
sitzen	0	0	0	45	0
treppen	0	0	0	0	4