

ABCD Study CT and Demographic Data Exploratory Data Analysis

Aidan Neher

2023-02-16

What is Exploratory Data Analysis?

Exploratory Data Analysis (EDA) is a structured approach for understanding your data that can be used for research question and hypothesis development. EDA's overall objective is to get insights to make better decisions. Sub-objectives include:

- Identify correlated variables.
- Identify and deal with outliers.
- Identify trends across time.
- Identify trends across space.
- Uncover patterns related to the response variable of interest.
- Create research questions to explore or hypotheses to test.
- Identify possible new data sources.

Set-Up Environment

The .RDS file loaded below was generated using the script “code/0_get_data.R”.

```
## -- Attaching packages ----- tidyverse 1.3.2 --
## v ggplot2 3.4.0      v purrr 0.3.5
## v tibble 3.1.8      v dplyr 1.0.10
## v tidyr 1.2.1      v stringr 1.4.1
## v readr 2.1.3      v forcats 0.5.2
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag() masks stats::lag()
```

About the Variables

- subjectkey is the subject's unique identifier.
- eventname is the data collection point for an observation (row of data). Note, interview_age is also available in months.
- Brain structure metrics cortical thickness (thick), sulcus depth (sulc), and surface area (area) are included. For more on the meaning of these metrics, see <https://doi-org.ezp2.lib.umn.edu/10.1007%2Fs00429-015-1177-6>

Explore the data

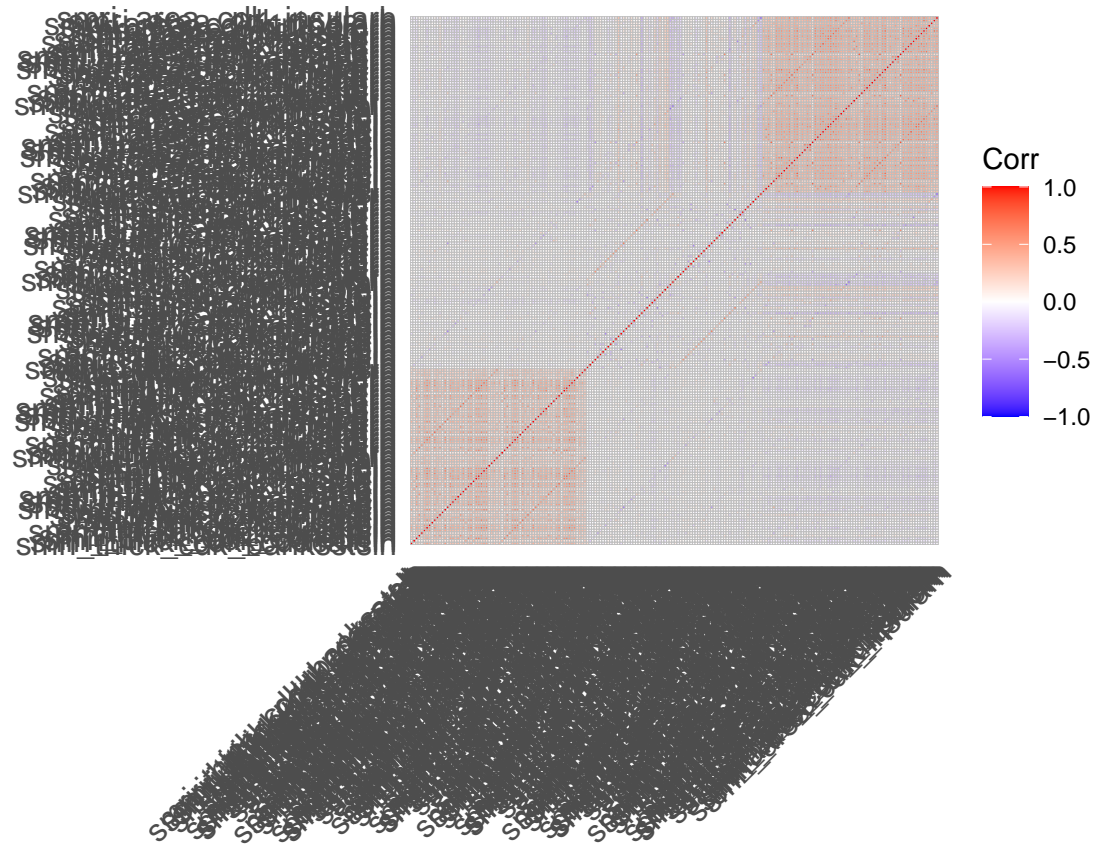
From <http://www.sthda.com/english/articles/31-principal-component-methods-in-r-practical-guide/118-principal-component-analysis-in-r-prcomp-vs-princomp/#prcomp-and-princomp-functions>:

“There are two general methods to perform PCA in R :

- Spectral decomposition which examines the covariances / correlations between variables
- Singular value decomposition which examines the covariances / correlations between individuals

The function `princomp()` uses the spectral decomposition approach. The functions `prcomp()` and `PCA()`[FactoMineR] use the singular value decomposition (SVD)."

```
##
## baseline_year_1_arm_1 2_year_follow_up_y_arm_1
## 11760 7827
## Welcome! Want to learn more? See two factoextra-related books at https://goo.gl/ve3WBa
## [1] 0
```



```
##
## Comp.1 Comp.2
## smri_thick_cdk_banksstslh 0.053848254 0.0947350334
## smri_thick_cdk_cdacatelh 0.041551692 0.0301376692
## smri_thick_cdk_cdmndfrlh 0.072889419 0.1178940488
## smri_thick_cdk_cuneuslh 0.030388977 0.1120217443
## smri_thick_cdk_ehinallh 0.021558316 0.0607375382
## smri_thick_cdk_fusiformlh 0.061794441 0.1265946102
## smri_thick_cdk_ifpllh 0.087197216 0.1366499999
## smri_thick_cdk_iftmlh 0.060609527 0.1305380324
## smri_thick_cdk_ihcatelh 0.028594714 0.0473859841
## smri_thick_cdk_locclh 0.062075941 0.1495217462
## smri_thick_cdk_lobfrlh 0.060896781 0.0991952091
## smri_thick_cdk_linguallh 0.015213219 0.1176795965
## smri_thick_cdk_mobfrlh 0.035931227 0.0391089999
## smri_thick_cdk_mdtmlh 0.077759782 0.1238745918
## smri_thick_cdk_parahpallh 0.035948086 0.0430618120
## smri_thick_cdk_paracnlh 0.070337810 0.1233858395
## smri_thick_cdk_parsopclh 0.070181550 0.1063225911
```

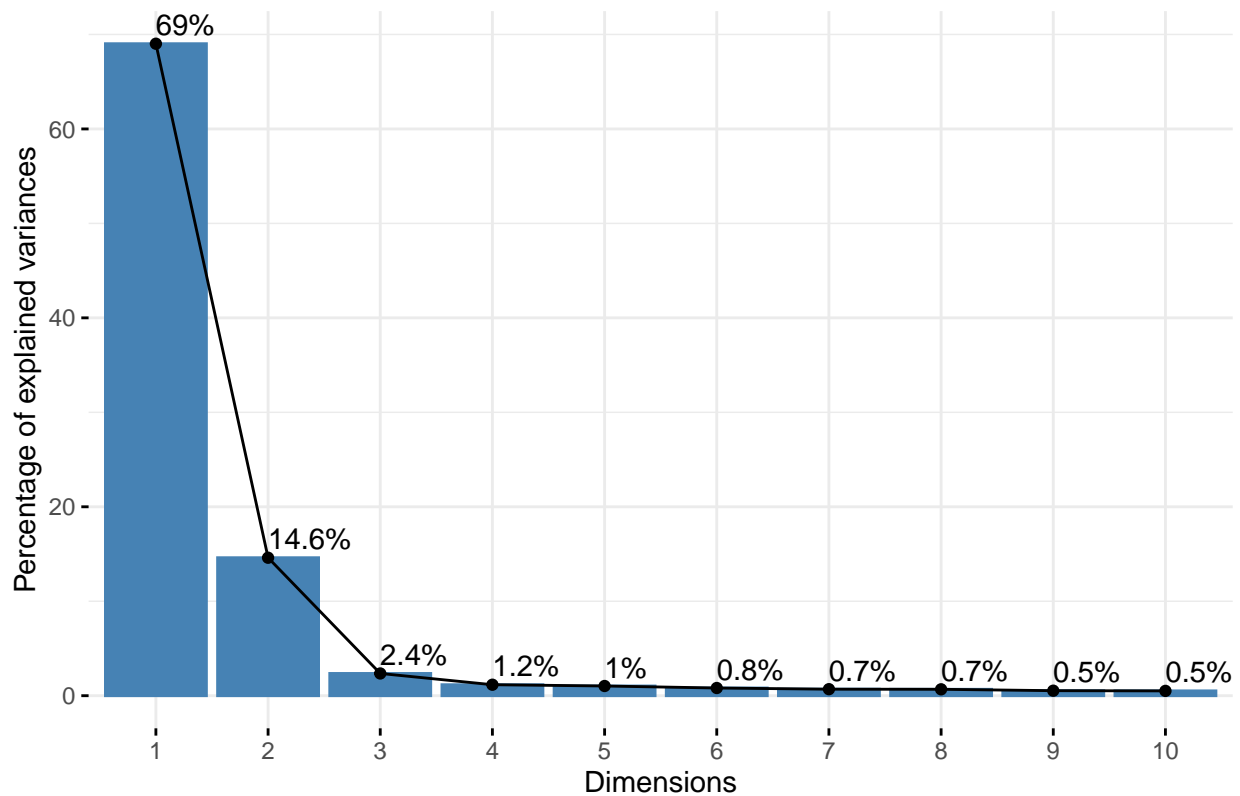
## smri_thick_cdk_parsobislh	0.074261168	0.0687439854
## smri_thick_cdk_parstgrislh	0.076107805	0.0939974294
## smri_thick_cdk_pericclh	0.030896336	0.1059695470
## smri_thick_cdk_postcnlh	0.055547292	0.1302873814
## smri_thick_cdk_ptcatelh	0.033436539	0.0628200394
## smri_thick_cdk_precnlh	0.071862128	0.1409131425
## smri_thick_cdk_pclh	0.077134123	0.1360541559
## smri_thick_cdk_rracatelh	0.050397074	0.0440857091
## smri_thick_cdk_rrmdfrlh	0.084729337	0.1278392607
## smri_thick_cdk_sufrlh	0.105709000	0.1145469221
## smri_thick_cdk_supllh	0.078217246	0.1489769375
## smri_thick_cdk_sutmlh	0.079896999	0.1254261377
## smri_thick_cdk_smlh	0.089579056	0.1304512387
## smri_thick_cdk_frpolelh	0.051939608	0.0397774103
## smri_thick_cdk_tmpolelh	0.050373604	0.0762806751
## smri_thick_cdk_trvtmlh	0.056471601	0.0895911102
## smri_thick_cdk_insulalh	0.024148583	0.0582163600
## smri_thick_cdk_banksstsrh	0.050670891	0.0931382641
## smri_thick_cdk_cdacaterh	0.028779386	0.0149004693
## smri_thick_cdk_cdmdfrrh	0.072579230	0.1165368050
## smri_thick_cdk_cuneusrh	0.035071614	0.1244479768
## smri_thick_cdk_ehinalrh	0.031726809	0.0564841064
## smri_thick_cdk_fusifformrh	0.060931518	0.1396019714
## smri_thick_cdk_ifplrh	0.088435304	0.1390498065
## smri_thick_cdk_iftmrh	0.065174720	0.1341581602
## smri_thick_cdk_ihcaterh	0.023785590	0.0361124151
## smri_thick_cdk_locrrh	0.057759951	0.1552814661
## smri_thick_cdk_lobfrrh	0.053851552	0.0840208703
## smri_thick_cdk_lingualrh	0.022625145	0.1248716739
## smri_thick_cdk_mobfrrh	0.044843288	0.0509151999
## smri_thick_cdk_mdtmrh	0.080926544	0.1280445817
## smri_thick_cdk_parahpalrh	0.051497457	0.0580329154
## smri_thick_cdk_paracnrh	0.062091537	0.1180451562
## smri_thick_cdk_parsopcrh	0.070772005	0.0963497720
## smri_thick_cdk_parsobisrh	0.076556831	0.0635046611
## smri_thick_cdk_parstgrisrh	0.078763356	0.0936429224
## smri_thick_cdk_periccrh	0.026983372	0.1056298778
## smri_thick_cdk_postcnrh	0.053762303	0.1148659333
## smri_thick_cdk_ptcaterh	0.028609716	0.0522504066
## smri_thick_cdk_precnrh	0.063450886	0.1252394223
## smri_thick_cdk_pcrh	0.075619782	0.1294433615
## smri_thick_cdk_rracaterh	0.035111348	0.0207325744
## smri_thick_cdk_rrmdfrrh	0.084437579	0.1113798662
## smri_thick_cdk_sufrh	0.098749833	0.1034662829
## smri_thick_cdk_suplrh	0.078667755	0.1465927511
## smri_thick_cdk_sutmrh	0.077414411	0.1300674778
## smri_thick_cdk_smrh	0.081427437	0.1297191528
## smri_thick_cdk_frpolerh	0.054713074	0.0364927793
## smri_thick_cdk_tmpolerh	0.052614955	0.0808731026
## smri_thick_cdk_trvtmrh	0.037108062	0.0773220550
## smri_thick_cdk_insularh	0.036538386	0.0718644544
## smri_sulc_cdk_banksstslh	0.023718648	-0.0122123792
## smri_sulc_cdk_cdacatelh	0.041572044	-0.0558540193
## smri_sulc_cdk_cdmdfrlh	0.027024303	-0.0227442720

## smri_sulc_cdk_cuneuslh	-0.007057796	0.0263166351
## smri_sulc_cdk_ehinallh	-0.009731221	0.0169728476
## smri_sulc_cdk_fusiformlh	-0.013613368	0.0132701626
## smri_sulc_cdk_ifpllh	0.001226533	-0.0026499016
## smri_sulc_cdk_iftmlh	-0.004852455	-0.0097662198
## smri_sulc_cdk_ihcatelh	0.029811746	-0.0057769342
## smri_sulc_cdk_locclh	-0.001258515	-0.0173906177
## smri_sulc_cdk_lobfrlh	0.013238930	0.0050677765
## smri_sulc_cdk_linguallh	-0.002063079	0.0277142905
## smri_sulc_cdk_mobfrlh	-0.051946604	0.0652750005
## smri_sulc_cdk_mdtmlh	0.008567422	-0.0056607434
## smri_sulc_cdk_parahpallh	-0.015459728	0.0066394562
## smri_sulc_cdk_paracnlh	-0.010935976	0.0092185550
## smri_sulc_cdk_parsopclh	0.003594092	-0.0133851522
## smri_sulc_cdk_parsobislh	-0.037232071	0.0060713509
## smri_sulc_cdk_parstgrislh	-0.004695053	0.0040539229
## smri_sulc_cdk_pericclh	-0.016743976	0.0259449116
## smri_sulc_cdk_postcnlh	-0.010117171	0.0233231205
## smri_sulc_cdk_ptcatelh	0.064896336	-0.0685698469
## smri_sulc_cdk_precnlh	0.003520786	-0.0029613856
## smri_sulc_cdk_pclh	-0.021693568	0.0222179068
## smri_sulc_cdk_rracatelh	0.026687447	-0.0372491490
## smri_sulc_cdk_rrmdfrlh	-0.001066652	-0.0341557532
## smri_sulc_cdk_sufrlh	-0.048474020	-0.0007411084
## smri_sulc_cdk_supllh	0.032590280	-0.0284897765
## smri_sulc_cdk_sutmlh	-0.049454102	0.0478746856
## smri_sulc_cdk_smlh	-0.007917346	0.0092396126
## smri_sulc_cdk_frpolelh	-0.051920037	0.0222906340
## smri_sulc_cdk_tmpolelh	-0.055407489	0.0269900508
## smri_sulc_cdk_trvtmlh	0.054067638	-0.0424815946
## smri_sulc_cdk_insulalh	0.042766868	-0.0561959225
## smri_sulc_cdk_banksstsrh	0.030034634	-0.0062883122
## smri_sulc_cdk_cdacaterh	0.039715919	-0.0577409441
## smri_sulc_cdk_cdmdfrrh	0.035835996	-0.0209065097
## smri_sulc_cdk_cuneusrh	0.009926659	0.0132841063
## smri_sulc_cdk_ehinalrh	-0.009605387	0.0231986290
## smri_sulc_cdk_fusiformrh	-0.008381148	0.0161382423
## smri_sulc_cdk_ifplrh	-0.007054781	0.0080053949
## smri_sulc_cdk_iftmrh	-0.003762619	-0.0062768525
## smri_sulc_cdk_ihcaterh	0.031708340	-0.0104704287
## smri_sulc_cdk_loccrh	-0.004517985	-0.0074518561
## smri_sulc_cdk_lobfrrh	0.010495085	0.0103147538
## smri_sulc_cdk_lingualrh	0.001415436	0.0234174380
## smri_sulc_cdk_mobfrrh	-0.061175690	0.0527070970
## smri_sulc_cdk_mdtmrh	0.003800504	-0.0025452871
## smri_sulc_cdk_parahpalrh	-0.025878634	0.0115091812
## smri_sulc_cdk_paracnrh	-0.004430326	0.0096735404
## smri_sulc_cdk_parsopcrh	-0.003089645	-0.0091152762
## smri_sulc_cdk_parsobisrh	-0.026941603	-0.0075607622
## smri_sulc_cdk_parstgrisrh	-0.013961189	0.0092942380
## smri_sulc_cdk_periccrh	-0.010461969	0.0164631205
## smri_sulc_cdk_postcnrh	-0.006754413	0.0147185607
## smri_sulc_cdk_ptcaterh	0.062164407	-0.0679640585
## smri_sulc_cdk_precnrh	0.005946012	-0.0051421160

## smri_sulc_cdk_pcrh	-0.024305799	0.0283028441
## smri_sulc_cdk_rracaterh	0.023460699	-0.0312929426
## smri_sulc_cdk_rrmdfrrh	0.002601724	-0.0337503113
## smri_sulc_cdk_sufrh	-0.055685304	0.0083661101
## smri_sulc_cdk_suplrh	0.032222860	-0.0288279873
## smri_sulc_cdk_sutmrh	-0.049693058	0.0520590198
## smri_sulc_cdk_smrh	0.014171510	-0.0084894840
## smri_sulc_cdk_frpolerh	-0.047021718	0.0201388590
## smri_sulc_cdk_tmpolerh	-0.047937969	0.0273260591
## smri_sulc_cdk_trvtmrh	0.048795573	-0.0485839379
## smri_sulc_cdk_insularh	0.051401689	-0.0612615546
## smri_area_cdk_banksstslh	-0.083251995	0.0381364209
## smri_area_cdk_cdacatelh	-0.072766515	0.0586950094
## smri_area_cdk_cdmdfrlh	-0.103629349	0.0407779645
## smri_area_cdk_cuneuslh	-0.078280408	0.0290102269
## smri_area_cdk_ehinallh	-0.047874898	0.0396243453
## smri_area_cdk_fusiformlh	-0.111804364	0.0689704668
## smri_area_cdk_ifpllh	-0.101644162	0.0403500840
## smri_area_cdk_iftmlh	-0.113588716	0.0541585804
## smri_area_cdk_ihcatelh	-0.091964512	0.0631472201
## smri_area_cdk_locclh	-0.103822624	0.0386764806
## smri_area_cdk_lobfrlh	-0.111093049	0.0922052733
## smri_area_cdk_lingualh	-0.078831787	0.0401811060
## smri_area_cdk_mobfrlh	-0.104100487	0.0691024334
## smri_area_cdk_mdtmlh	-0.121395588	0.0450925203
## smri_area_cdk_parahpallh	-0.070970269	0.0547116314
## smri_area_cdk_paracnlh	-0.094049008	0.0482912484
## smri_area_cdk_parsopclh	-0.086393379	0.0453975700
## smri_area_cdk_parsobislh	-0.104231895	0.0656647597
## smri_area_cdk_parstgrislh	-0.092330067	0.0434990365
## smri_area_cdk_pericclh	-0.057158843	0.0211008918
## smri_area_cdk_postcnlh	-0.121028308	0.0474061018
## smri_area_cdk_ptcatelh	-0.093143026	0.0717226833
## smri_area_cdk_precnlh	-0.121082677	0.0423907400
## smri_area_cdk_pclh	-0.109436374	0.0757517616
## smri_area_cdk_rracatelh	-0.100637521	0.0802636442
## smri_area_cdk_rrmdfrlh	-0.116502358	0.0551244583
## smri_area_cdk_sufrlh	-0.129112442	0.0574925226
## smri_area_cdk_supllh	-0.104293584	0.0410112750
## smri_area_cdk_sutmlh	-0.125548076	0.0523610172
## smri_area_cdk_smlh	-0.104731550	0.0492279848
## smri_area_cdk_frpolelh	-0.094742271	0.0578439817
## smri_area_cdk_tmpolelh	-0.089072540	0.0472101707
## smri_area_cdk_trvtmlh	-0.086834290	0.0439343915
## smri_area_cdk_insulalh	-0.105003454	0.0872085394
## smri_area_cdk_banksstsrh	-0.089840446	0.0461089260
## smri_area_cdk_cdacaterh	-0.074563403	0.0568371559
## smri_area_cdk_cdmdfrrh	-0.100956074	0.0403968658
## smri_area_cdk_cuneusrh	-0.084921296	0.0328500708
## smri_area_cdk_ehinalrh	-0.051824324	0.0405849285
## smri_area_cdk_fusiformrh	-0.113615833	0.0689133212
## smri_area_cdk_ifplrh	-0.106521357	0.0481902764
## smri_area_cdk_iftmrh	-0.116771604	0.0575655775
## smri_area_cdk_ihcaterh	-0.088778846	0.0578090826

## smri_area_cdk_loccrh	-0.102884565	0.0450099512
## smri_area_cdk_lobfrrh	-0.113846478	0.0698109388
## smri_area_cdk_lingualrh	-0.077468585	0.0386287841
## smri_area_cdk_mobfrrh	-0.110122804	0.0813242515
## smri_area_cdk_mdtmrh	-0.127904953	0.0532913113
## smri_area_cdk_parahpalrh	-0.073133258	0.0543449480
## smri_area_cdk_paracnrh	-0.092851788	0.0448772617
## smri_area_cdk_parsopcrh	-0.090897787	0.0459664392
## smri_area_cdk_parsobisrh	-0.098863453	0.0642817166
## smri_area_cdk_parstgrish	-0.089512628	0.0371349820
## smri_area_cdk_periccrh	-0.061082304	0.0212071479
## smri_area_cdk_postcnrh	-0.116082960	0.0461890234
## smri_area_cdk_ptcaterh	-0.097141562	0.0673724950
## smri_area_cdk_precnrh	-0.117726879	0.0420559854
## smri_area_cdk_pcrh	-0.112529296	0.0784606659
## smri_area_cdk_rracaterh	-0.089030276	0.0762333119
## smri_area_cdk_rrmdfrrh	-0.112482143	0.0529244081
## smri_area_cdk_sufrh	-0.124920117	0.0599357100
## smri_area_cdk_suplrh	-0.105800495	0.0438925981
## smri_area_cdk_sutmrh	-0.127301132	0.0545461858
## smri_area_cdk_smrh	-0.106229556	0.0494802347
## smri_area_cdk_frpolerh	-0.089963392	0.0583024870
## smri_area_cdk_tmpolerh	-0.082137656	0.0413110398
## smri_area_cdk_trvtmrh	-0.099401765	0.0462680668
## smri_area_cdk_insularh	-0.105658163	0.0803035151

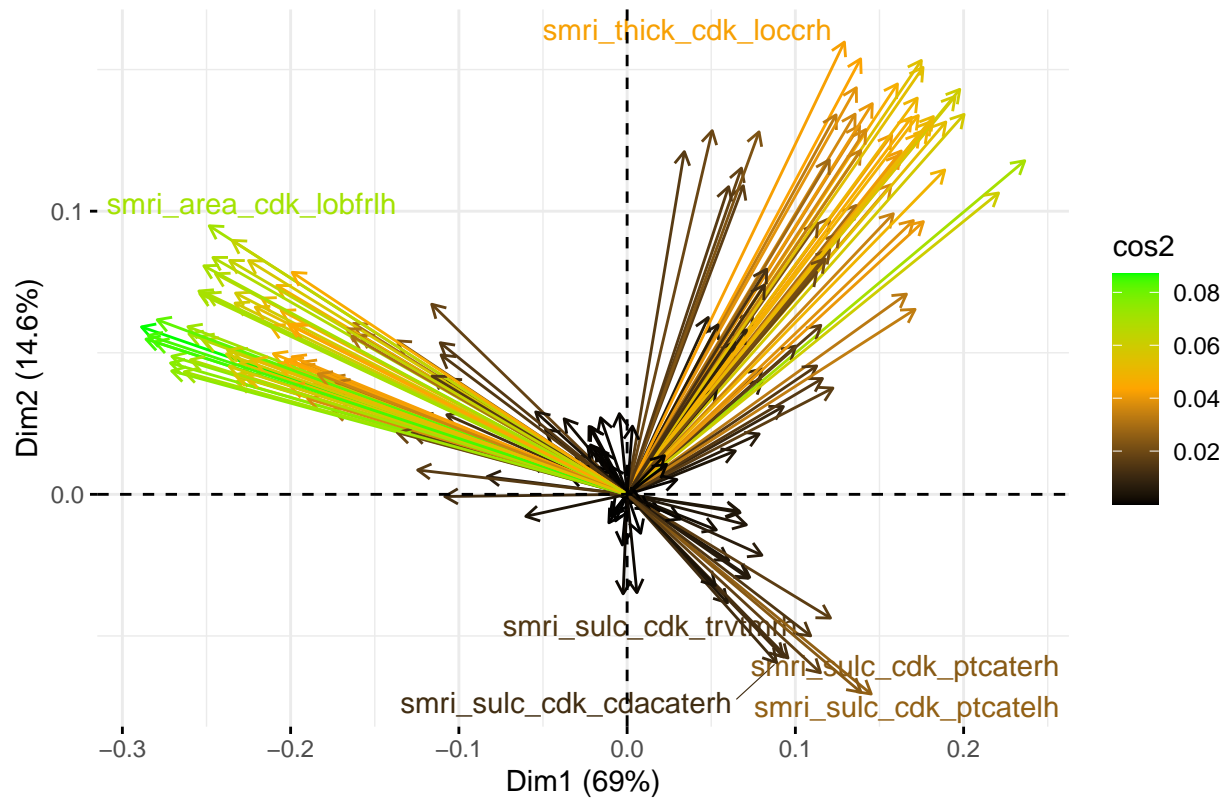
Scree plot



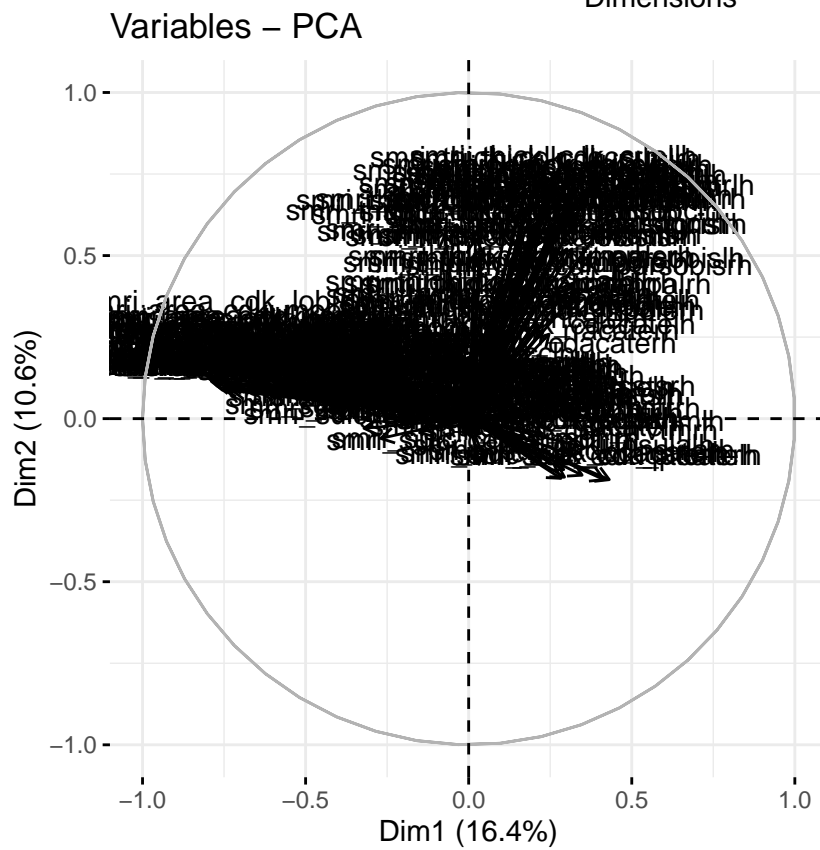
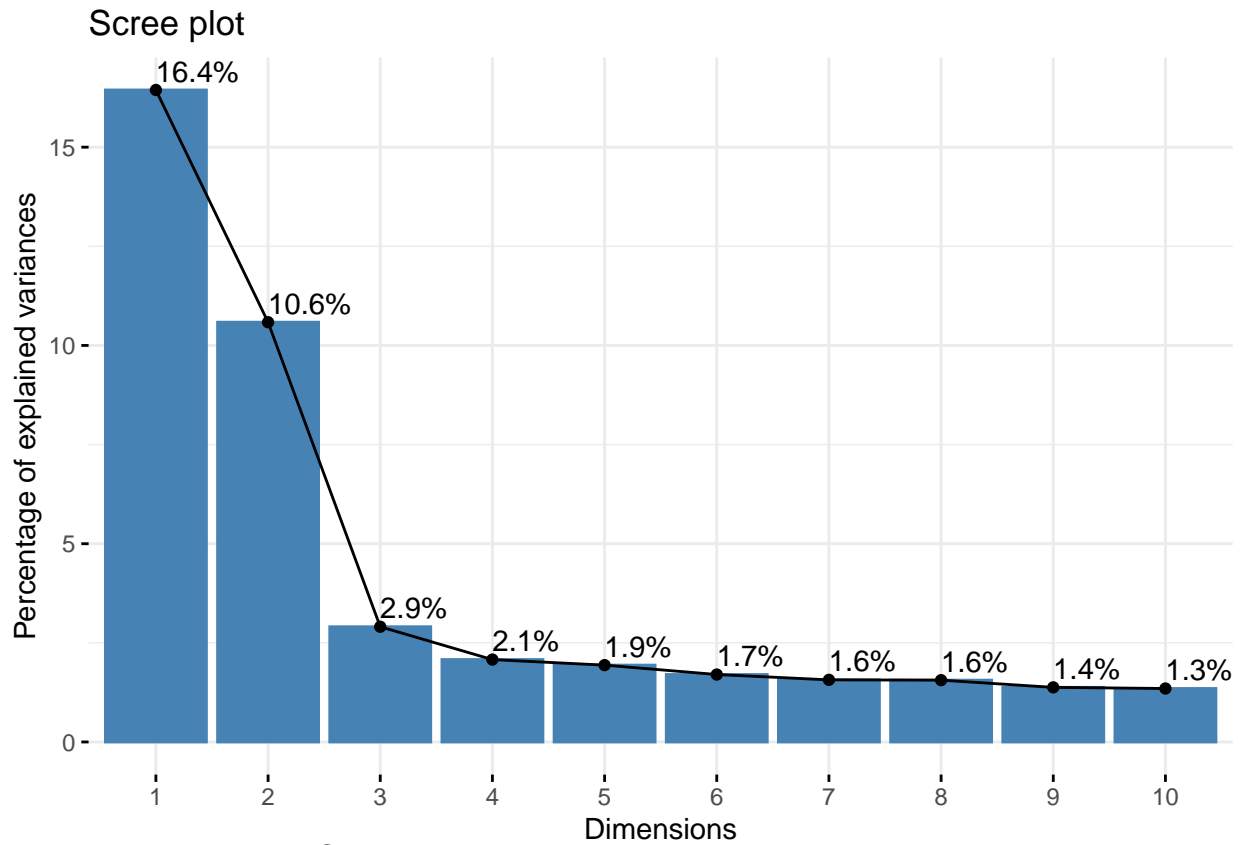
7

increasing max.overlaps

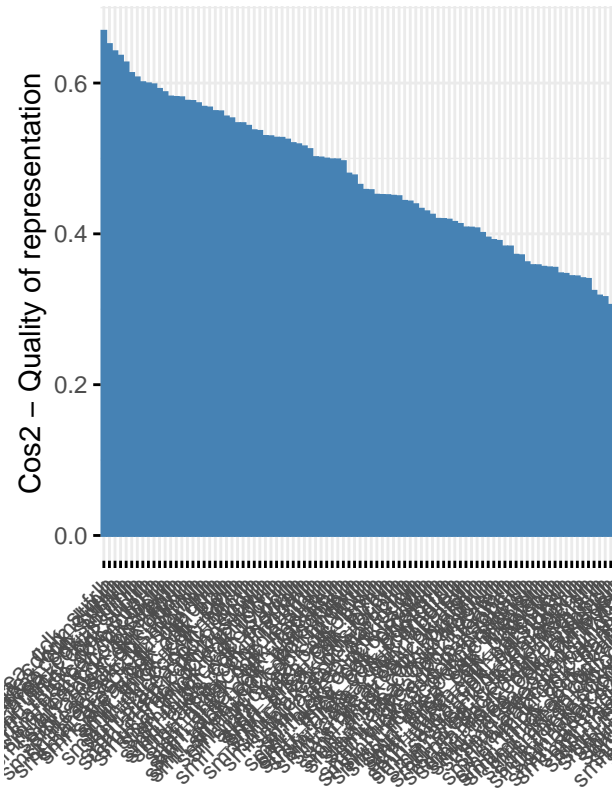
Variables – PCA



NULL

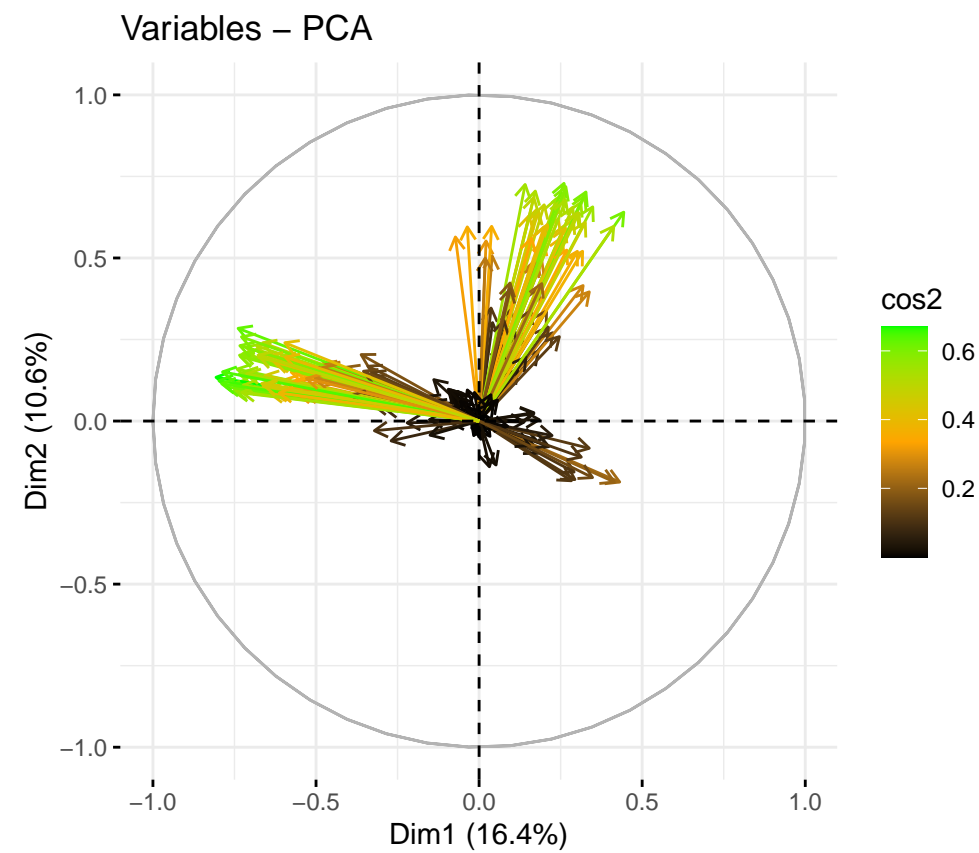


Cos2 of variables to Dim-1-2



Warning: ggrepel: 204 unlabeled data points (too many overlaps). Consider

increasing max.overlaps



Questions

- There's more than one demo_data observation per subjectkey, why?
- What other covariates are required?