ABCD Study CT and Demographic Data Exploratory Data Analysis

Aidan Neher

2023-02-23

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) 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

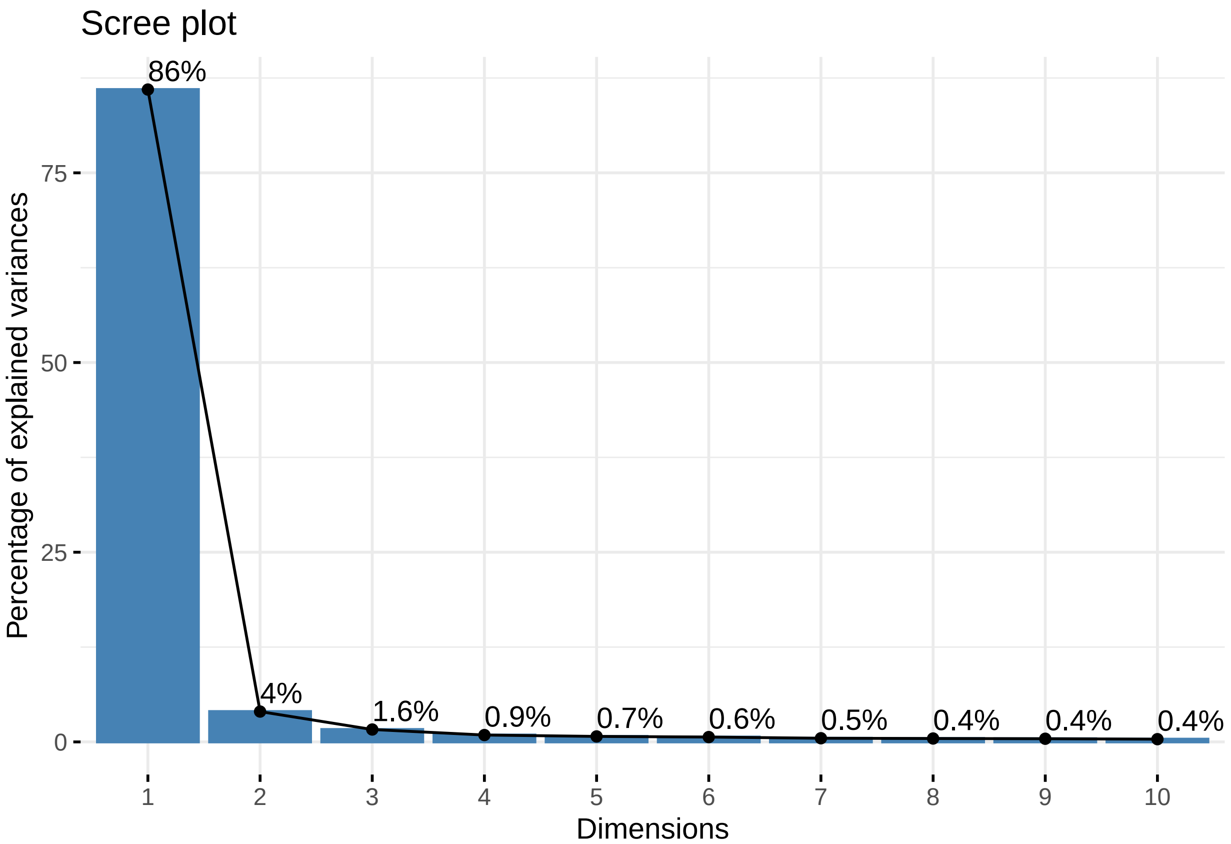
The function princomp() uses the spectral decomposition approach. The functions prcomp() and PCA()[FactoMineR] use the singular value decomposition (SVD)."

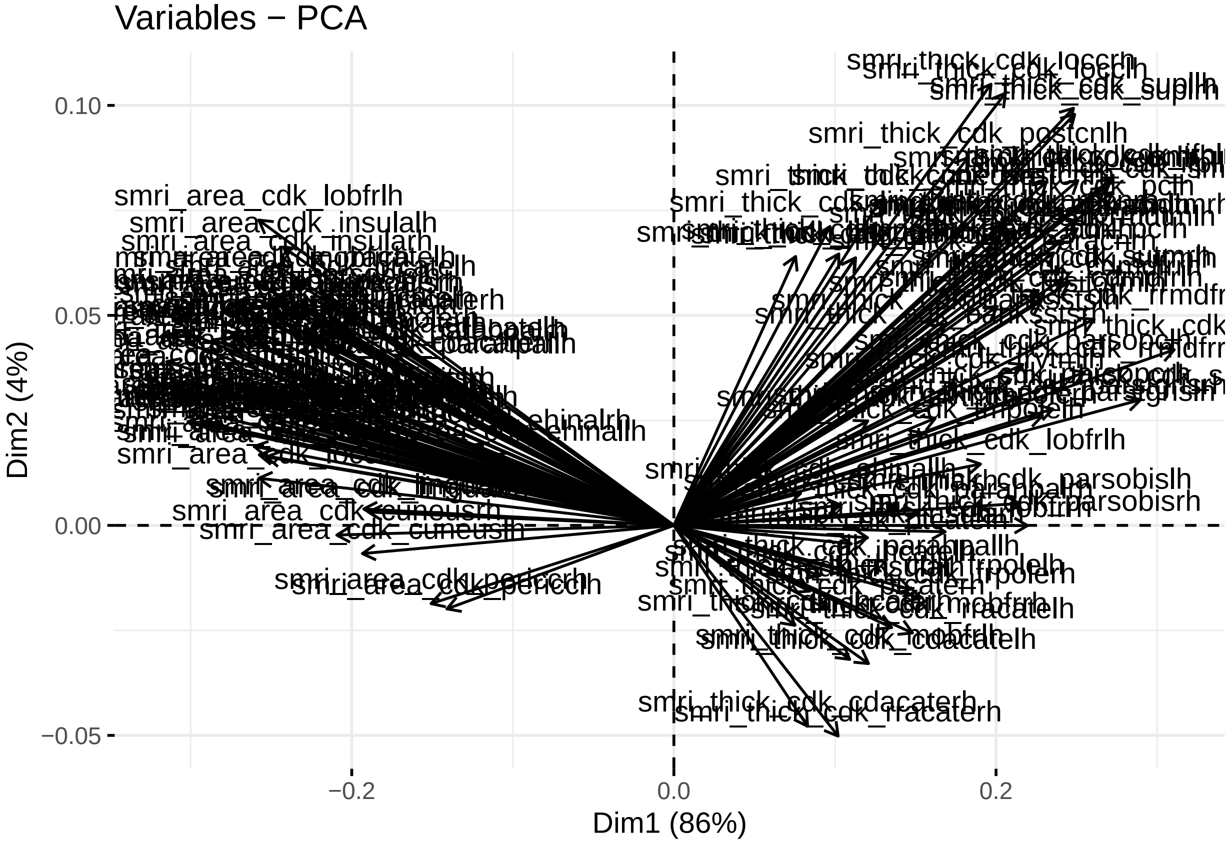
```
## 
##    baseline_year_1_arm_1 2_year_follow_up_y_arm_1 
##                    11760                     7827
```

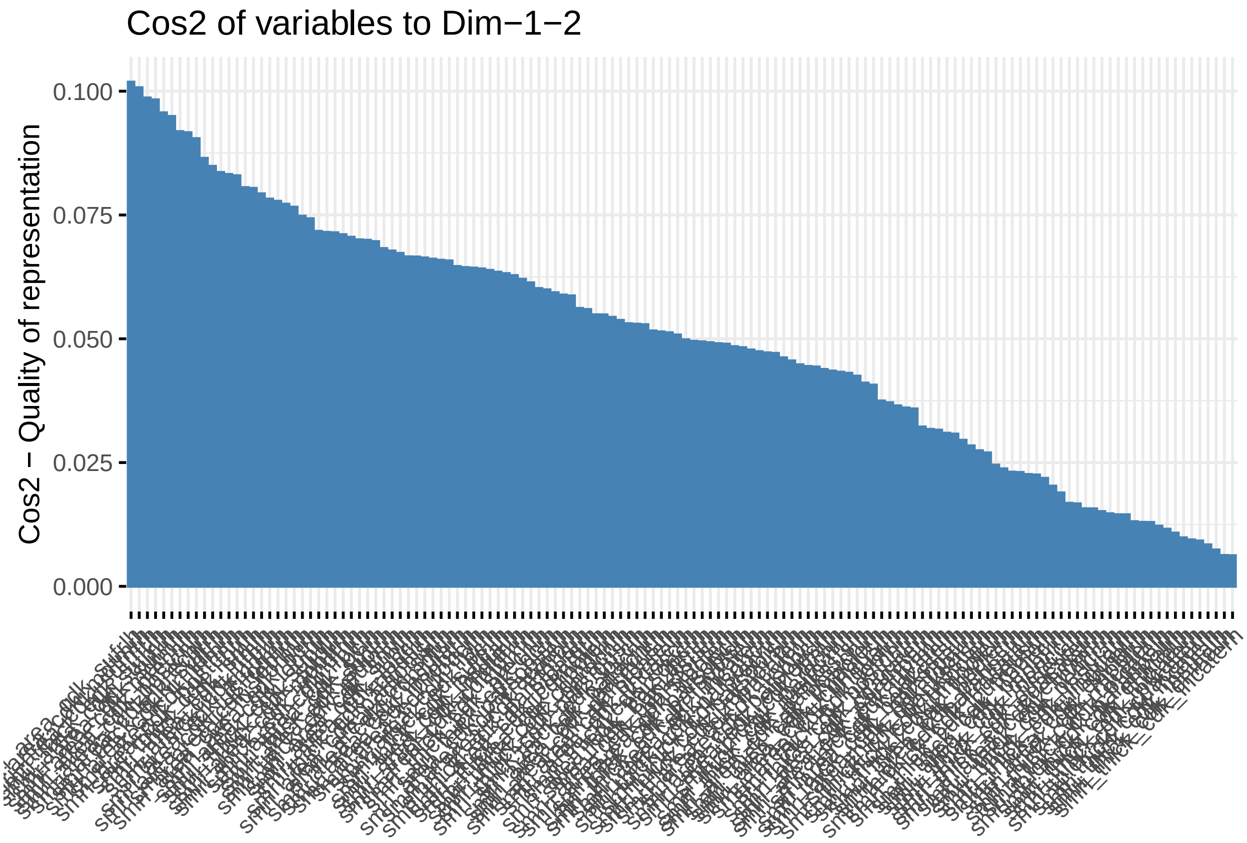
```
## Welcome! Want to learn more? See two factoextra-related books at https://goo.gl/ve3WBa
```

```
## [1] 0
```

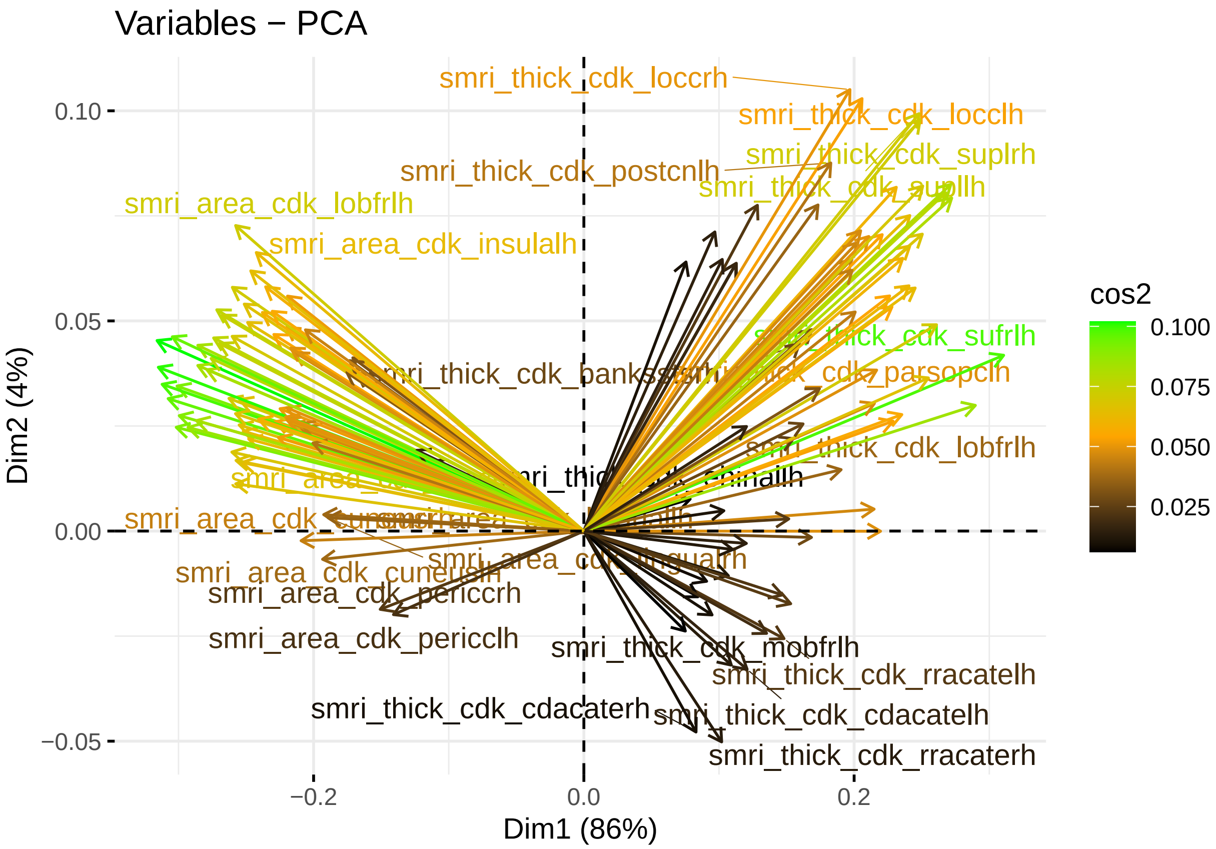
```
##                                 Comp.1        Comp.2
## smri_thick_cdk_bankssstslh  0.06625518  8.725898e-02
## smri_thick_cdk_cdacatelh    0.04745296 -5.991249e-02
## smri_thick_cdk_cdmdfrlh     0.08955769  9.712527e-02
## smri_thick_cdk_cuneuslh     0.04413288  1.157824e-01
## smri_thick_cdk_ehinallh     0.03090478  1.388150e-02
## smri_thick_cdk_fusiformlh   0.07886650  9.490028e-02
## smri_thick_cdk_ifpllh       0.10518480  1.465818e-01
## smri_thick_cdk_iftmlh       0.07786942  1.177056e-01
## smri_thick_cdk_ihcatelh     0.03568816 -2.176728e-02
## smri_thick_cdk_locclh       0.08083702  1.873481e-01
## smri_thick_cdk_lobfrlh      0.07477291  2.660233e-02
## smri_thick_cdk_linguallh    0.02967834  1.165807e-01
## smri_thick_cdk_mobfrlh      0.04286902 -5.810143e-02
## smri_thick_cdk_mdtmlh       0.09471701  1.234793e-01
## smri_thick_cdk_parahpallh   0.04200549 -1.937588e-02
## smri_thick_cdk_paracnlh     0.08670675  1.284155e-01
## smri_thick_cdk_parsopclh    0.08509736  6.974297e-02
## smri_thick_cdk_parsobislh   0.08435655  9.496219e-03
## smri_thick_cdk_parstgrislh  0.08987492  4.759859e-02
## smri_thick_cdk_pericclh     0.04431071  1.160356e-01
## smri_thick_cdk_postcnlh     0.07176371  1.594504e-01
## smri_thick_cdk_ptcatelh     0.04328125 -8.006222e-03
## smri_thick_cdk_precnlh      0.09080641  1.489300e-01
## smri_thick_cdk_pclh         0.09473222  1.366822e-01
## smri_thick_cdk_rracatelh    0.05818145 -4.662266e-02
## smri_thick_cdk_rrmdfrlh     0.10248388  8.937420e-02
## smri_thick_cdk_sufrlh       0.12199996  7.615331e-02
## smri_thick_cdk_supllh       0.09743695  1.808285e-01
## smri_thick_cdk_sutmlh       0.09629015  1.053049e-01
## smri_thick_cdk_smlh         0.10690614  1.442738e-01
## smri_thick_cdk_frpolelh     0.05771306 -2.769021e-02
## smri_thick_cdk_tmpolelh     0.06085725  4.004324e-02
## smri_thick_cdk_trvtmlh      0.06847943  6.174183e-02
## smri_thick_cdk_insulalh     0.03313666 -2.849160e-02
## smri_thick_cdk_bankssstsrh  0.06261008  8.112469e-02
## smri_thick_cdk_cdacaterh    0.03259387 -8.707324e-02
## smri_thick_cdk_cdmdfrrh     0.08886833  1.018681e-01
## smri_thick_cdk_cuneusrh     0.05042952  1.411348e-01
## smri_thick_cdk_ehinalrh     0.04070193  8.808797e-03
## smri_thick_cdk_fusiformrh   0.07936969  1.248439e-01
## smri_thick_cdk_ifplrh       0.10663684  1.501452e-01
## smri_thick_cdk_iftmrh       0.08281932  1.276220e-01
## smri_thick_cdk_ihcaterh     0.02948976 -4.333352e-02
## smri_thick_cdk_loccrh       0.07732509  1.912566e-01
## smri_thick_cdk_lobfrrh      0.06615811 -2.766386e-03
## smri_thick_cdk_lingualrh    0.03807327  1.295971e-01
## smri_thick_cdk_mobfrrh      0.05311335 -4.432638e-02
## smri_thick_cdk_mdtmrh       0.09837146  1.286349e-01
## smri_thick_cdk_parahpalrh   0.05955657  5.261711e-03
## smri_thick_cdk_paracnrh     0.07795897  1.129978e-01
## smri_thick_cdk_parsopcrh    0.08445059  5.530963e-02
## smri_thick_cdk_parsobisrh   0.08623234 -8.193278e-05
## smri_thick_cdk_parstgrisrh  0.09241259  5.046184e-02
## smri_thick_cdk_periccrh     0.04022474  1.176866e-01
## smri_thick_cdk_postcnrh     0.06804966  1.412861e-01
## smri_thick_cdk_ptcaterh     0.03726937 -3.634485e-02
## smri_thick_cdk_precnrh      0.08044186  1.301003e-01
## smri_thick_cdk_pcrh         0.09266887  1.181475e-01
## smri_thick_cdk_rracaterh    0.04004670 -9.135830e-02
## smri_thick_cdk_rrmdfrrh     0.10013254  6.634164e-02
## smri_thick_cdk_sufrrh       0.11377982  5.444726e-02
## smri_thick_cdk_suplrh       0.09772031  1.781865e-01
## smri_thick_cdk_sutmrh       0.09450172  1.062433e-01
## smri_thick_cdk_smrh         0.09848852  1.494784e-01
## smri_thick_cdk_frpolerh     0.06012274 -3.149786e-02
## smri_thick_cdk_tmpolerh     0.06371092  4.641212e-02
## smri_thick_cdk_trvtmrh      0.04727889  4.528632e-02
## smri_thick_cdk_insularh     0.04723117 -5.348018e-03
## smri_area_cdk_banksstslh   -0.08033042  4.520592e-02
## smri_area_cdk_cdacatelh    -0.06715374  7.506177e-02
## smri_area_cdk_cdmdfrlh     -0.10019756  4.597656e-02
## smri_area_cdk_cuneuslh     -0.07595260 -1.223322e-02
## smri_area_cdk_ehinallh     -0.04443291  3.030206e-02
## smri_area_cdk_fusiformlh   -0.10594020  8.129145e-02
## smri_area_cdk_ifpllh       -0.09900993  4.328386e-02
## smri_area_cdk_iftmlh       -0.10954351  6.990419e-02
## smri_area_cdk_ihcatelh     -0.08642887  8.819984e-02
## smri_area_cdk_locclh       -0.10119901  2.043876e-02
## smri_area_cdk_lobfrlh      -0.10121158  1.323607e-01
## smri_area_cdk_linguallh    -0.07562028  7.118952e-03
## smri_area_cdk_mobfrlh      -0.09772757  9.036839e-02
## smri_area_cdk_mdtmlh       -0.11824336  6.281704e-02
## smri_area_cdk_parahpallh   -0.06583939  6.820493e-02
## smri_area_cdk_paracnlh     -0.08970609  4.738532e-02
## smri_area_cdk_parsopclh    -0.08210549  4.460266e-02
## smri_area_cdk_parsobislh   -0.09870651  9.832238e-02
## smri_area_cdk_parstgrislh  -0.08829895  5.300558e-02
## smri_area_cdk_pericclh     -0.05534264 -3.620337e-02
## smri_area_cdk_postcnlh     -0.11760161  4.999433e-02
## smri_area_cdk_ptcatelh     -0.08612988  1.017994e-01
## smri_area_cdk_precnlh      -0.11848031  4.495131e-02
## smri_area_cdk_pclh         -0.10222589  8.435836e-02
## smri_area_cdk_rracatelh    -0.09248881  1.058131e-01
## smri_area_cdk_rrmdfrlh     -0.11221016  8.046841e-02
## smri_area_cdk_sufrlh       -0.12404722  8.253084e-02
## smri_area_cdk_supllh       -0.10110256  3.059959e-02
## smri_area_cdk_sutmlh       -0.12082389  5.738851e-02
## smri_area_cdk_smlh         -0.10004037  5.696510e-02
## smri_area_cdk_frpolelh     -0.09010898  8.516515e-02
## smri_area_cdk_tmpolelh     -0.08522241  4.659528e-02
## smri_area_cdk_trvtmlh      -0.08215763  4.655359e-02
## smri_area_cdk_insulalh     -0.09520782  1.206549e-01
## smri_area_cdk_banksstsrh   -0.08635848  5.344327e-02
## smri_area_cdk_cdacaterh    -0.06888343  6.876221e-02
## smri_area_cdk_cdmdfrrh     -0.09743096  3.992116e-02
## smri_area_cdk_cuneusrh     -0.08219641 -4.186676e-03
## smri_area_cdk_ehinalrh     -0.04856657  3.496880e-02
## smri_area_cdk_fusiformrh   -0.10760776  8.366488e-02
## smri_area_cdk_ifplrh       -0.10315971  5.747916e-02
## smri_area_cdk_iftmrh       -0.11221105  7.172902e-02
## smri_area_cdk_ihcaterh     -0.08386823  7.727606e-02
## smri_area_cdk_loccrh       -0.09936582  2.968089e-02
## smri_area_cdk_lobfrrh      -0.10669621  9.585576e-02
## smri_area_cdk_lingualrh    -0.07435117  5.765738e-03
## smri_area_cdk_mobfrrh      -0.10216136  1.055956e-01
## smri_area_cdk_mdtmrh       -0.12368351  7.096221e-02
## smri_area_cdk_parahpalrh   -0.06809894  7.384378e-02
## smri_area_cdk_paracnrh     -0.08878071  4.049181e-02
## smri_area_cdk_parsopcrh    -0.08671576  4.945976e-02
## smri_area_cdk_parsobisrh   -0.09343402  9.469834e-02
## smri_area_cdk_parstgrisrh  -0.08632640  4.823609e-02
## smri_area_cdk_periccrh     -0.05920669 -3.391929e-02
## smri_area_cdk_postcnrh     -0.11262752  4.786273e-02
## smri_area_cdk_ptcaterh     -0.09056850  9.521519e-02
## smri_area_cdk_precnrh      -0.11506149  4.490698e-02
## smri_area_cdk_pcrh         -0.10509258  9.354396e-02
## smri_area_cdk_rracaterh    -0.08092727  8.716658e-02
## smri_area_cdk_rrmdfrrh     -0.10832513  7.478968e-02
## smri_area_cdk_sufrrh       -0.11961226  8.415938e-02
## smri_area_cdk_suplrh       -0.10231581  3.422757e-02
## smri_area_cdk_sutmrh       -0.12257846  6.359012e-02
## smri_area_cdk_smrh         -0.10125252  5.088284e-02
## smri_area_cdk_frpolerh     -0.08516839  7.928717e-02
## smri_area_cdk_tmpolerh     -0.07875887  3.810496e-02
## smri_area_cdk_trvtmrh      -0.09454924  4.852220e-02
## smri_area_cdk_insularh     -0.09681344  1.127197e-01
```


Scree plot
Percentage of explained variances
0
25
50
75
86%
4%
1.6%
0.9%
0.7%
0.6%
0.5%
0.4%
0.4%
0.4%
1
2
3
4
5
6
7
8
9
10
Dimensions

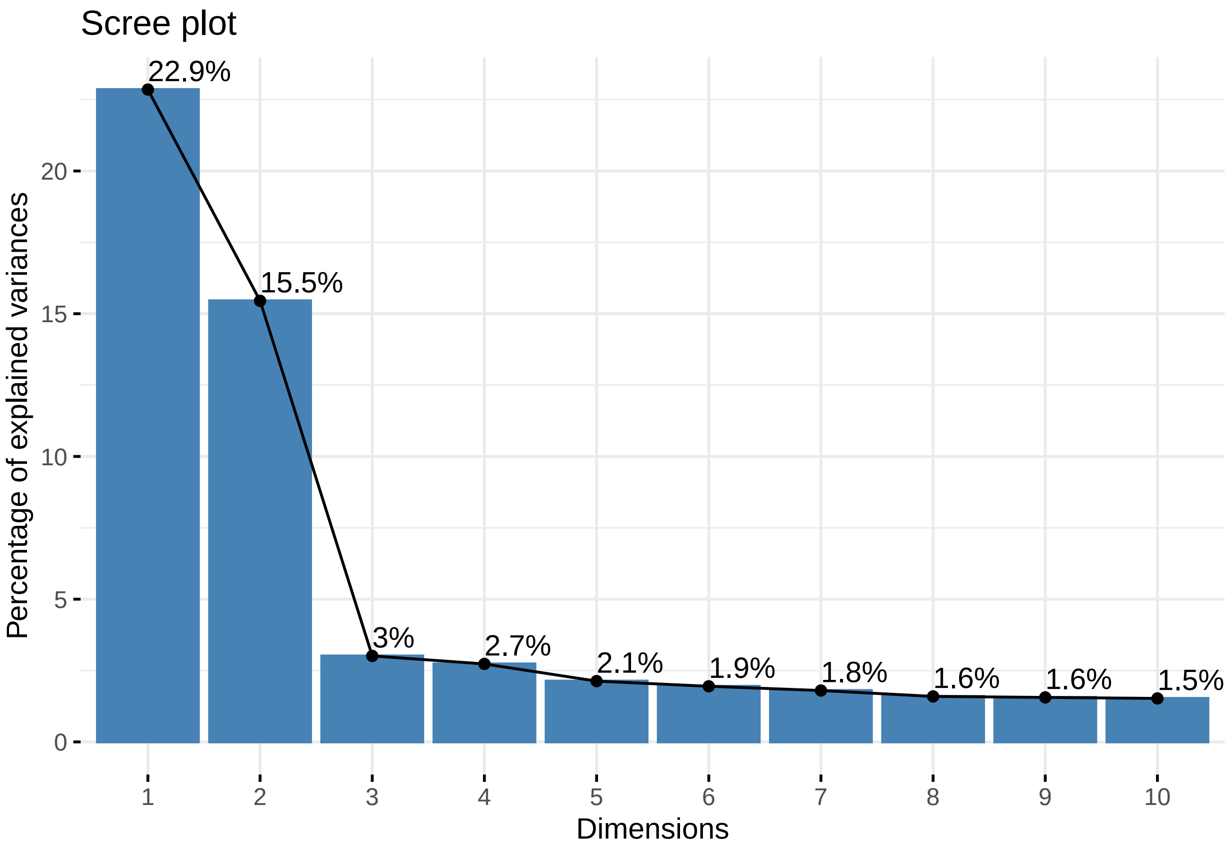

Variables − PCA
Dim2 (4%)
Dim1 (86%)
0.10
0.05
0.00
−0.05
−0.2
0.0
0.2
smri_thick_cdk_loccrh
smri_thick_cdk_loccrlh
smri_thick_cdk_supllh
smri_thick_cdk_suplrh
smri_thick_cdk_postcnlh
smri_thick_cdk_pclh
smri_thick_cdk_rrmdfr
smri_thick_cdk_parsopcrh
smri_thick_cdk_lobfrlh
smri_thick_cdk_parsobislh
smri_thick_cdk_parsobisrh
smri_thick_cdk_ehinallh
smri_thick_cdk_parahpallh
smri_thick_cdk_frpolelh
smri_thick_cdk_mobfrlh
smri_thick_cdk_racatelh
smri_thick_cdk_cdacatelh
smri_thick_cdk_cdacaterh
smri_thick_cdk_racaterh
smri_area_cdk_lobfrlh
smri_area_cdk_insulalh
smri_area_cdk_insularh
smri_area_cdk_ehinalh
smri_area_cdk_ehinalrh
smri_area_cdk_lingualh
smri_area_cdk_cuneusrh
smri_area_cdk_cuneuslh
smri_area_cdk_pericclh
smri_area_cdk_periccrh

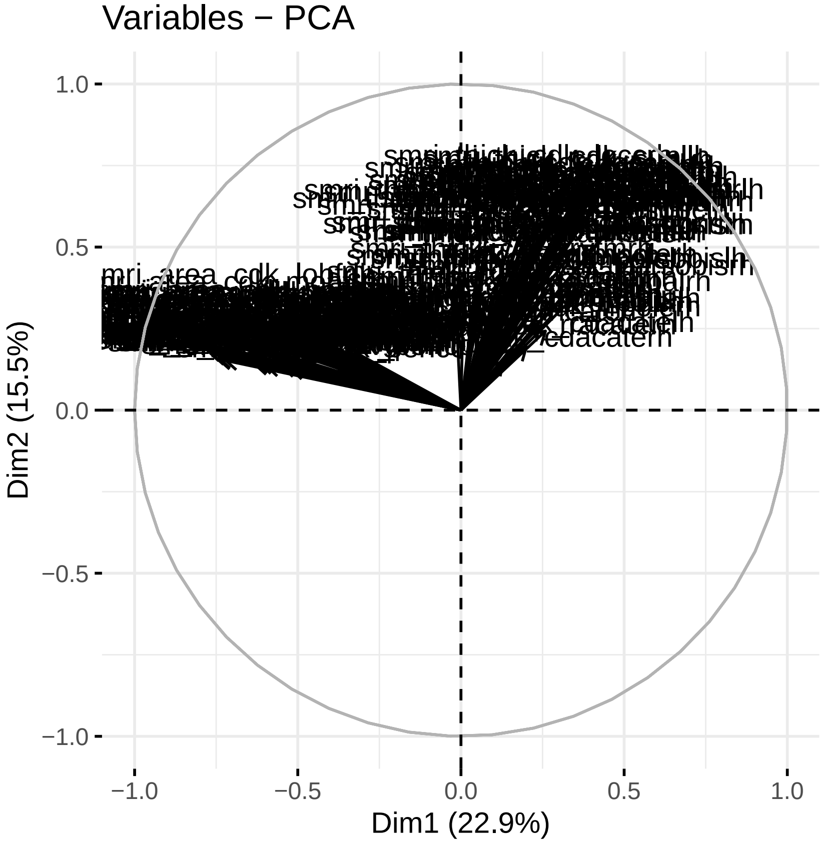


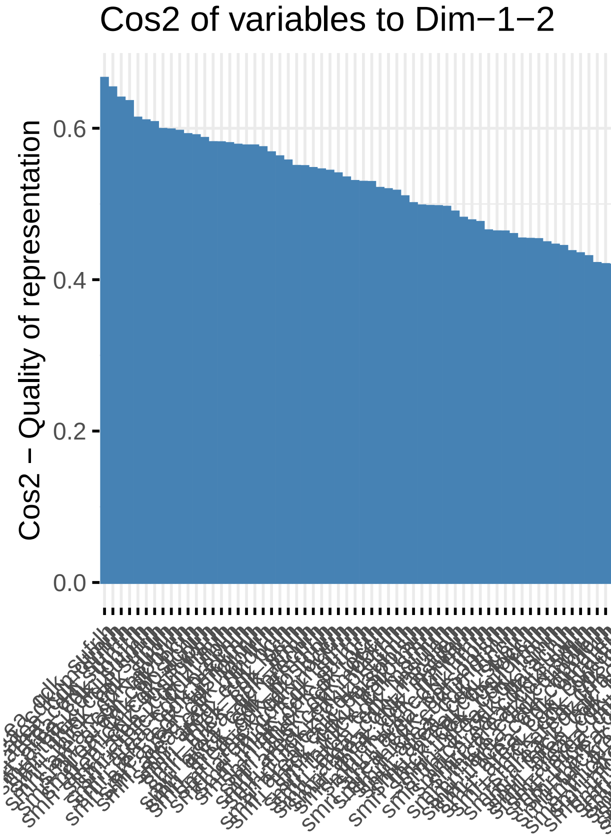
```
## Warning: ggrepel: 112 unlabeled data points (too many overlaps). Consider
## increasing max.overlaps
```


Variables − PCA
smri_thick_cdk_loccrh
smri_thick_cdk_locclh
smri_thick_cdk_suplrh
smri_thick_cdk_postcnlh
smri_thick_cdk_supllh
smri_area_cdk_lobfrlh
smri_area_cdk_insulalh
smri_thick_cdk_sufrlh
smri_thick_cdk_bankssts
smri_thick_cdk_parsopclh
smri_thick_cdk_lobfrlh
smri_area_cdk_cuneusrh
smri_area_cdk_lingualrh
smri_area_cdk_cuneuslh
smri_area_cdk_periccrh
smri_area_cdk_pericclh
smri_thick_cdk_mobfrlh
smri_thick_cdk_rracatelh
smri_thick_cdk_cdacaterh
smri_thick_cdk_cdacatelh
smri_thick_cdk_rracaterh
Dim2 (4%)
Dim1 (86%)
0.10
0.05
0.00
−0.05
−0.2
0.0
0.2
cos2
0.100
0.075
0.050
0.025

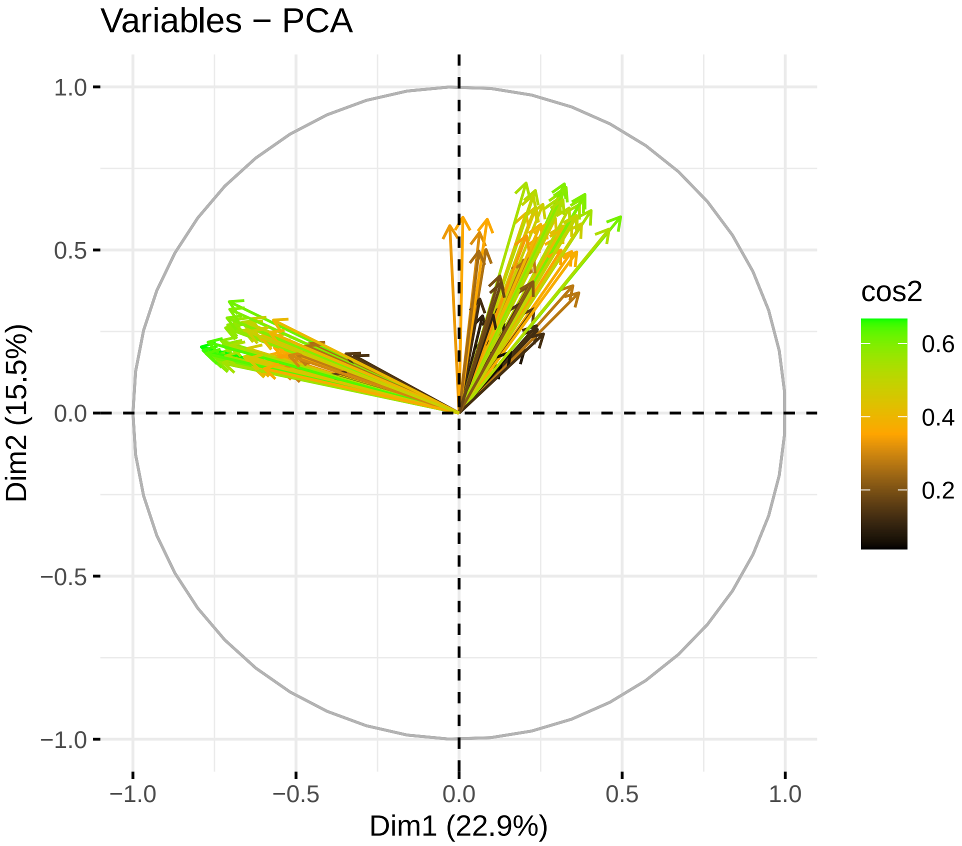
```
## NULL
```







```
## Warning: ggrepel: 136 unlabeled data points (too many overlaps). Consider
## increasing max.overlaps
```



```
## Warning: In lm.fit(x, y, offset = offset, singular.ok = singular.ok, ...) :
##  extra argument 'family' will be disregarded

##
## Call:
## lm(formula = outcome_si ~ ., data = classification_data, family = binomial)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.23298 -0.09958 -0.08080 -0.06259  0.99949
##
## Coefficients:
##                         Estimate Std. Error t value Pr(>|t|)
## (Intercept)           -0.0479139  0.1704187  -0.281  0.77860
## sexM                   0.0199268  0.0063840   3.121  0.00180 **
## interview_age         -0.0000104  0.0003772  -0.028  0.97801
## demo_comb_income_v22   0.0402188  0.0208160   1.932  0.05337 .
## demo_comb_income_v23  -0.0084748  0.0235227  -0.360  0.71865
## demo_comb_income_v24  -0.0024994  0.0201809  -0.124  0.90144
## demo_comb_income_v25   0.0095579  0.0194624   0.491  0.62337
## demo_comb_income_v26   0.0260332  0.0188992   1.377  0.16839
## demo_comb_income_v27   0.0165603  0.0186273   0.889  0.37401
## demo_comb_income_v28   0.0158443  0.0191767   0.826  0.40870
## demo_comb_income_v29   0.0055481  0.0189761   0.292  0.77001
## demo_comb_income_v210 -0.0066995  0.0205092  -0.327  0.74393
## demo_ethn_v22          0.0207575  0.0083980   2.472  0.01346 *
## demo_prnt_marital_v22  -0.0158013  0.0315690  -0.501  0.61671
## demo_prnt_marital_v23   0.0284504  0.0101581   2.801  0.00511 **
## demo_prnt_marital_v24  -0.0053194  0.0154692  -0.344  0.73095
## demo_prnt_marital_v25   0.0036616  0.0112107   0.327  0.74396
## demo_prnt_marital_v26   0.0349790  0.0132065   2.649  0.00809 **
## race_white              0.0026524  0.0106756   0.248  0.80379
## race_black              0.0021034  0.0108631   0.194  0.84647
## race_native             0.0451434  0.0153343   2.944  0.00325 **
## race_pacific_islander   0.0158049  0.0348715   0.453  0.65039
## race_asian              0.0025061  0.0120287   0.208  0.83497
## race_other              0.0018092  0.0142997   0.127  0.89933
## demo_prnt_highest_ed4   0.0059755  0.2056656   0.029  0.97682
## demo_prnt_highest_ed5   0.0597432  0.3252438   0.184  0.85426
## demo_prnt_highest_ed6   0.0948786  0.1727550   0.549  0.58287
## demo_prnt_highest_ed7   0.0217868  0.1943031   0.112  0.91072
## demo_prnt_highest_ed8   0.0529826  0.1737182   0.305  0.76038
## demo_prnt_highest_ed9   0.0453623  0.1670663   0.272  0.78599
## demo_prnt_highest_ed10  0.0530679  0.1664305   0.319  0.74984
## demo_prnt_highest_ed11  0.0951138  0.1649798   0.577  0.56428
## demo_prnt_highest_ed12  0.0675117  0.1646903   0.410  0.68187
## demo_prnt_highest_ed13  0.0888978  0.1628509   0.546  0.58516
## demo_prnt_highest_ed14  0.0701055  0.1633810   0.429  0.66787
## demo_prnt_highest_ed15  0.0897837  0.1626488   0.552  0.58095
## demo_prnt_highest_ed16  0.0907340  0.1628041   0.557  0.57732
## demo_prnt_highest_ed17  0.1145121  0.1629262   0.703  0.48217
## demo_prnt_highest_ed18  0.0891506  0.1626968   0.548  0.58373
## demo_prnt_highest_ed19  0.0878521  0.1627359   0.540  0.58932
## demo_prnt_highest_ed20  0.1008520  0.1631773   0.618  0.53656
## demo_prnt_highest_ed21  0.1028745  0.1630889   0.631  0.52819
## PC1                     0.0002771  0.0005747   0.482  0.62972
## PC2                    -0.0016336  0.0006551  -2.494  0.01265 *
## PC3                     0.0002606  0.0013686   0.190  0.84900
## PC4                    -0.0032922  0.0014525  -2.267  0.02343 *
## PC5                    -0.0030718  0.0016961  -1.811  0.07016 .
## PC6                    -0.0005862  0.0017268  -0.339  0.73428
## PC7                     0.0031464  0.0017891   1.759  0.07868 .
## PC8                    -0.0006794  0.0018935  -0.359  0.71977
## PC9                     0.0024827  0.0019654   1.263  0.20653
## PC10                    0.0020032  0.0019239   1.041  0.29779
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2808 on 10322 degrees of freedom
##   (1386 observations deleted due to missingness)
## Multiple R-squared:  0.009497,   Adjusted R-squared:  0.004603
## F-statistic: 1.941 on 51 and 10322 DF,  p-value: 6.965e-05
```

```
##
## Call:
## lm(formula = outcome_internalizing_score ~ ., data = regression_data)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -8.710 -3.732 -1.609  2.073 45.300
```

```
##
## Coefficients:
##                         Estimate Std. Error t value Pr(>|t|)
## (Intercept)             2.220507   3.292131   0.674 0.500015
## sexM                    0.228323   0.122883   1.858 0.063189 .
## interview_age           0.009223   0.007262   1.270 0.204087
## demo_comb_income_v22   -0.088178   0.400628  -0.220 0.825797
## demo_comb_income_v23   -0.001158   0.452580  -0.003 0.997958
## demo_comb_income_v24   -0.082435   0.389009  -0.212 0.832181
## demo_comb_income_v25   -0.325977   0.375390  -0.868 0.385212
## demo_comb_income_v26   -0.796779   0.364217  -2.188 0.028717 *
## demo_comb_income_v27   -1.206769   0.359305  -3.359 0.000786 ***
## demo_comb_income_v28   -1.659512   0.369852  -4.487 7.30e-06 ***
## demo_comb_income_v29   -2.018336   0.365856  -5.517 3.54e-08 ***
## demo_comb_income_v210  -2.608701   0.395496  -6.596 4.43e-11 ***
## demo_ethn_v22           0.253697   0.161528   1.571 0.116306
## demo_prnt_marital_v22   1.141591   0.603028   1.893 0.058372 .
## demo_prnt_marital_v23   0.375202   0.195721   1.917 0.055262 .
## demo_prnt_marital_v24   0.710747   0.298530   2.381 0.017292 *
## demo_prnt_marital_v25   0.348366   0.215732   1.615 0.106382
## demo_prnt_marital_v26   0.512366   0.254612   2.012 0.044210 *
## race_white              1.636374   0.205360   7.968 1.78e-15 ***
## race_black              0.035235   0.208939   0.169 0.866087
## race_native             0.483911   0.295443   1.638 0.101470
## race_pacific_islander  -0.176951   0.668932  -0.265 0.791379
## race_asian              0.102643   0.231427   0.444 0.657396
## race_other              1.062321   0.274915   3.864 0.000112 ***
## demo_prnt_highest_ed4  -0.341626   3.974112  -0.086 0.931498
## demo_prnt_highest_ed5   9.539833   6.284824   1.518 0.129066
## demo_prnt_highest_ed6   0.943504   3.338228   0.283 0.777461
## demo_prnt_highest_ed7   1.734818   3.754595   0.462 0.644054
## demo_prnt_highest_ed8   2.728861   3.356829   0.813 0.416278
## demo_prnt_highest_ed9   1.149969   3.226730   0.356 0.721557
## demo_prnt_highest_ed10  0.544359   3.214728   0.169 0.865538
## demo_prnt_highest_ed11  0.608225   3.187582   0.191 0.848677
## demo_prnt_highest_ed12  0.091056   3.181301   0.029 0.977166
## demo_prnt_highest_ed13  0.360597   3.146769   0.115 0.908770
## demo_prnt_highest_ed14  0.544492   3.156947   0.172 0.863068
## demo_prnt_highest_ed15  1.500629   3.142891   0.477 0.633039
## demo_prnt_highest_ed16  1.591244   3.145857   0.506 0.612992
## demo_prnt_highest_ed17  1.695386   3.148246   0.539 0.590231
## demo_prnt_highest_ed18  1.431699   3.143818   0.455 0.648830
## demo_prnt_highest_ed19  1.416982   3.144579   0.451 0.652279
## demo_prnt_highest_ed20  1.164309   3.153031   0.369 0.711937
## demo_prnt_highest_ed21  1.594926   3.151402   0.506 0.612797
## PC1                     0.040694   0.011072   3.675 0.000239 ***
## PC2                    -0.023211   0.012621  -1.839 0.065927 .
## PC3                     0.037480   0.026355   1.422 0.155017
## PC4                    -0.043514   0.027967  -1.556 0.119760
## PC5                     0.023682   0.032702   0.724 0.468978
## PC6                     0.000424   0.033219   0.013 0.989817
## PC7                     0.007232   0.034441   0.210 0.833682
## PC8                    -0.022266   0.036431  -0.611 0.541099
## PC9                     0.050588   0.037897   1.335 0.181943
## PC10                    -0.005547   0.037047  -0.150 0.880993
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.427 on 10382 degrees of freedom
##   (1326 observations deleted due to missingness)
## Multiple R-squared:  0.03109,    Adjusted R-squared:  0.02633
## F-statistic: 6.531 on 51 and 10382 DF,  p-value: < 2.2e-16
```

Questions

- There's more than one demo_data observation per subjectkey, why?
- What other covariates are required?
- Regarding time point determination for DSEM, do we want to use interview age or eventname?