Understanding Behavioural Data with PCA in R

Rubén Guerrero Ramírez

Data Scientist @

smart motors

boosting knowledge

smartmotors.org

Agenda

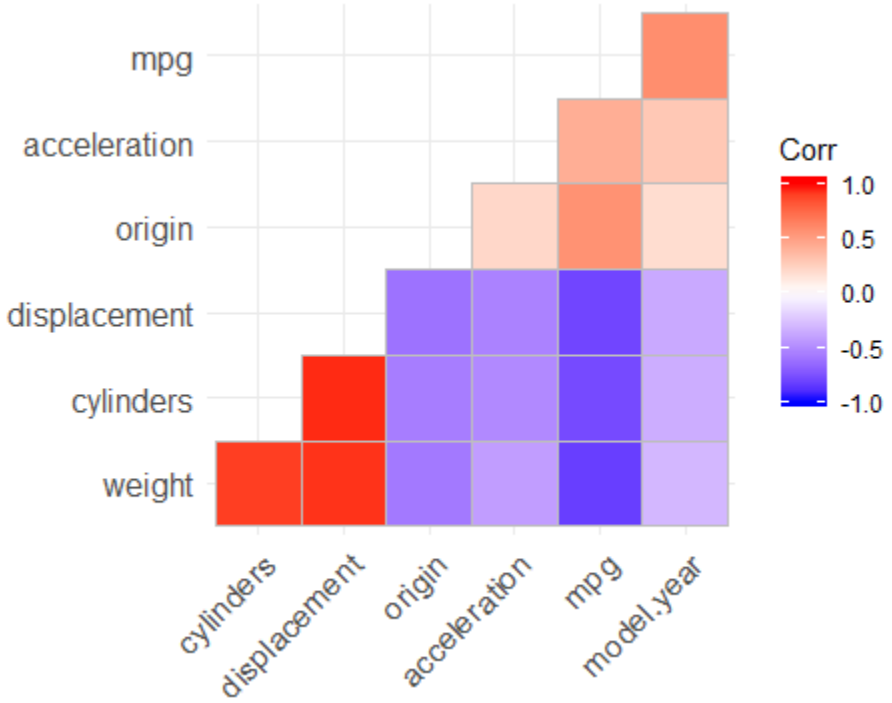
- (Very short) Introduction to PCA
- Why PCA?
- Constrains
- R example
- Other techniques

(Very short) Introduction to PCA

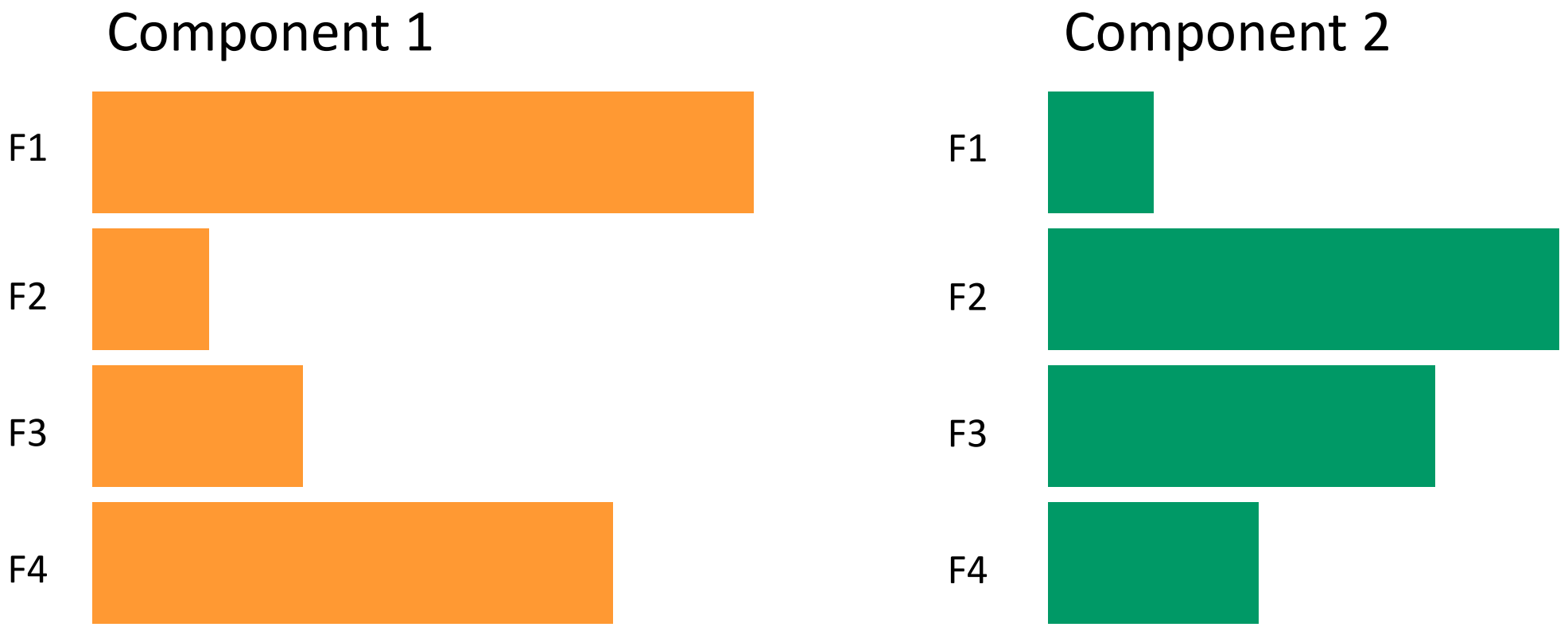
	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
1	21.5	4	120.1	97	3.70	2.465	20.01	1	0	3	1
2	21.4	6	258.0	110	3.08	3.215	19.44	1	0	3	1
3	19.2	8	400.0	175	3.08	3.845	17.05	0	0	3	2
4	18.7	8	360.0	175	3.15	3.440	17.02	0	0	3	2
5	18.1	6	225.0	105	2.76	3.460	20.22	1	0	3	1
6	17.3	8	275.8	180	3.07	3.730	17.60	0	0	3	3
7	16.4	8	275.8	180	3.07	4.070	17.40	0	0	3	3
8	15.5	8	318.0	150	2.76	3.520	16.87	0	0	3	2
9	15.2	8	275.8	180	3.07	3.780	18.00	0	0	3	3
10	15.2	8	304.0	150	3.15	3.435	17.30	0	0	3	2
11	14.7	8	440.0	230	3.23	5.345	17.42	0	0	3	4
12	14.3	8	360.0	245	3.21	3.570	15.84	0	0	3	4
13	13.3	8	350.0	245	3.73	3.840	15.41	0	0	3	4
14	10.4	8	472.0	205	2.93	5.250	17.98	0	0	3	4
15	10.4	8	460.0	215	3.00	5.424	17.82	0	0	3	4
16	33.9	4	71.1	65	4.22	1.835	19.90	1	1	4	1
17	32.4	4	78.7	66	4.08	2.200	19.47	1	1	4	1
18	30.4	4	75.7	52	4.93	1.615	18.52	1	1	4	2
19	27.3	4	79.0	66	4.08	1.935	18.90	1	1	4	1

(Very short) Introduction to PCA

	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
1	21.5	4	120.1	97	3.70	2.465	20.01	1	0	3	1
2	21.4	6	258.0	110	3.08	3.215	19.44	1	0	3	1
3	19.2	8	400.0	175	3.08	3.845	17.05	0	0	3	2
4	18.7	8	360.0	175	3.15	3.440	17.02	0	0	3	2
5	18.1	6	225.0	105	2.76	3.460	20.22	1	0	3	1
6	17.3	8	275.8	180	3.07	3.730	17.60	0	0	3	3
7	16.4	8	275.8	180	3.07	4.070	17.40	0	0	3	3
8	15.5	8	318.0	150	2.76	3.520	16.87	0	0	3	2
9	15.2	8	275.8	180	3.07	3.780	18.00	0	0	3	3
10	15.2	8	304.0	150	3.15	3.435	17.30	0	0	3	2
11	14.7	8	440.0	230	3.23	5.345	17.42	0	0	3	4
12	14.3	8	360.0	245	3.21	3.570	15.84	0	0	3	4
13	13.3	8	350.0	245	3.73	3.840	15.41	0	0	3	4
14	10.4	8	472.0	205	2.93	5.250	17.98	0	0	3	4
15	10.4	8	460.0	215	3.00	5.424	17.82	0	0	3	4
16	33.9	4	71.1	65	4.22	1.835	19.90	1	1	4	1
17	32.4	4	78.7	66	4.08	2.200	19.47	1	1	4	1
18	30.4	4	75.7	52	4.93	1.615	18.52	1	1	4	2
19	27.3	4	79.0	66	4.08	1.935	18.90	1	1	4	1



(Very short) Introduction to PCA



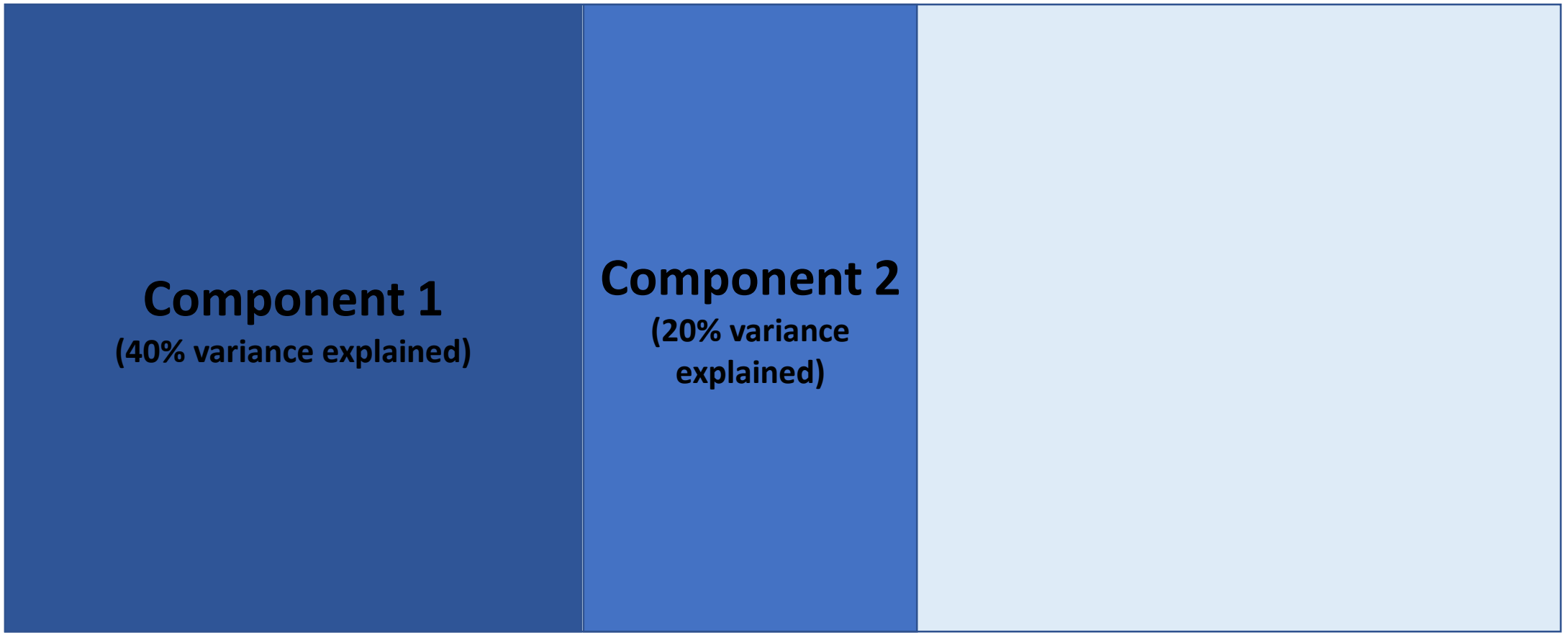
(Very short) Introduction to PCA

Total variance of the dataset

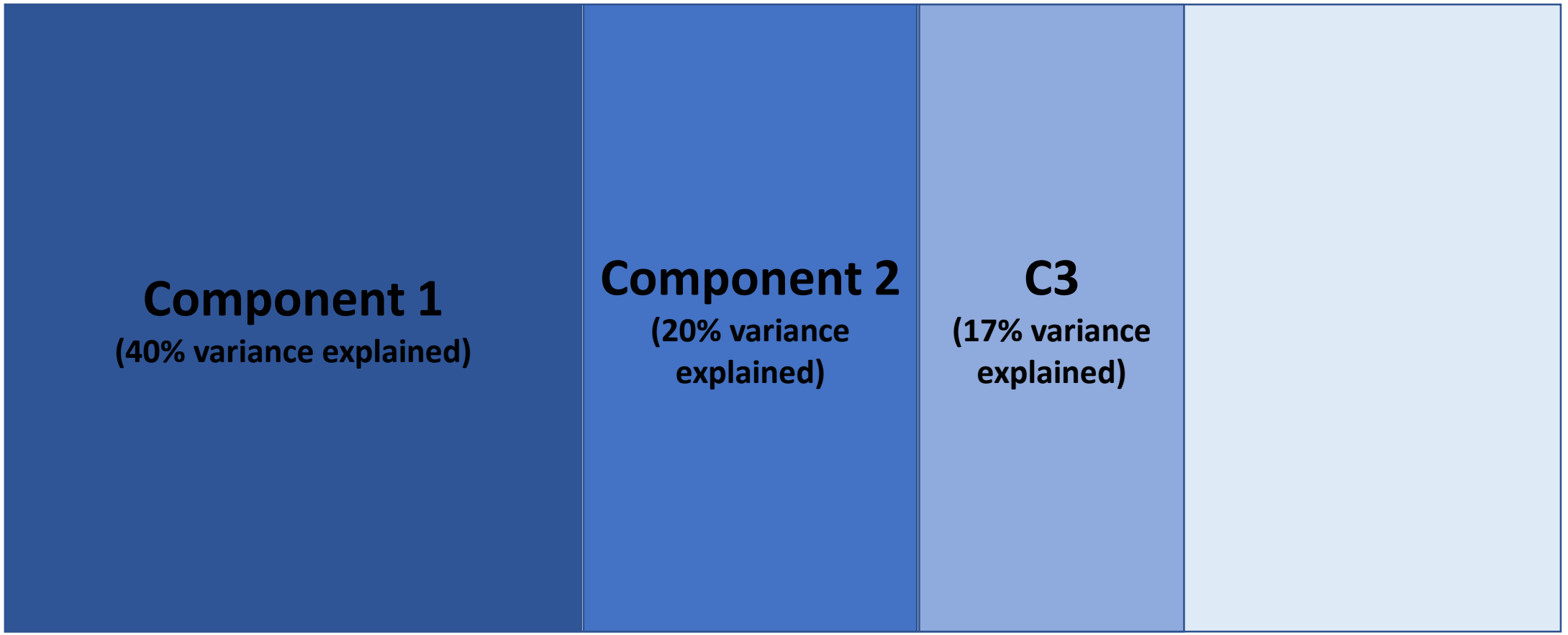
(Very short) Introduction to PCA

Component 1
(40% variance explained)

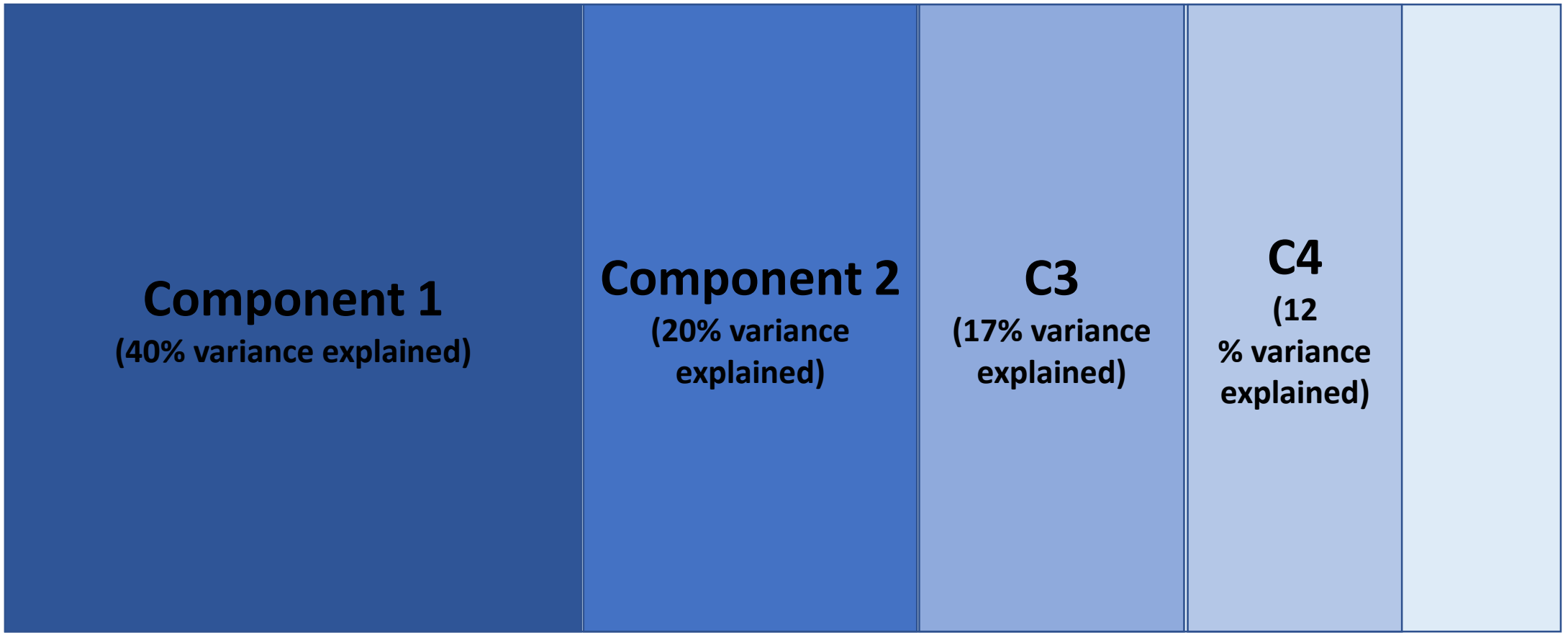
(Very short) Introduction to PCA



(Very short) Introduction to PCA

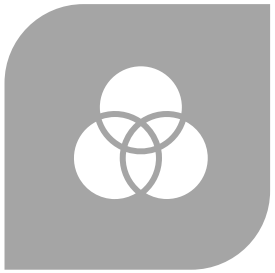


(Very short) Introduction to PCA



Why PCA?

REDUCES MODELING TIME



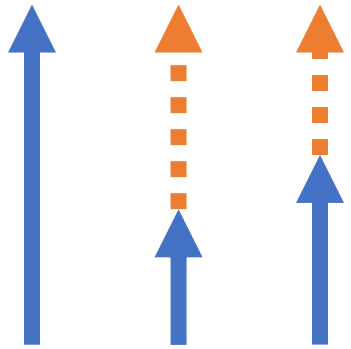
GROUPS CORRELATED FEATURES

EXTRACTS MEANINGFUL COMPONENTS ON QUESTIONNAIRES

ANOMALY DETECTION

Constrains

Missing values not allowed in dataset



All features must be scaled / normalized

R Example

Dataset:

- Humor Styles Questionnaire (http://www.humorstyles.com/)
- 1071 participants on 32 questions

R Example

Dataset:

- Humor Styles Questionnaire (http://www.humorstyles.com/)
- 1071 participants on 32 questions

Totally Disagree

Totally Agree

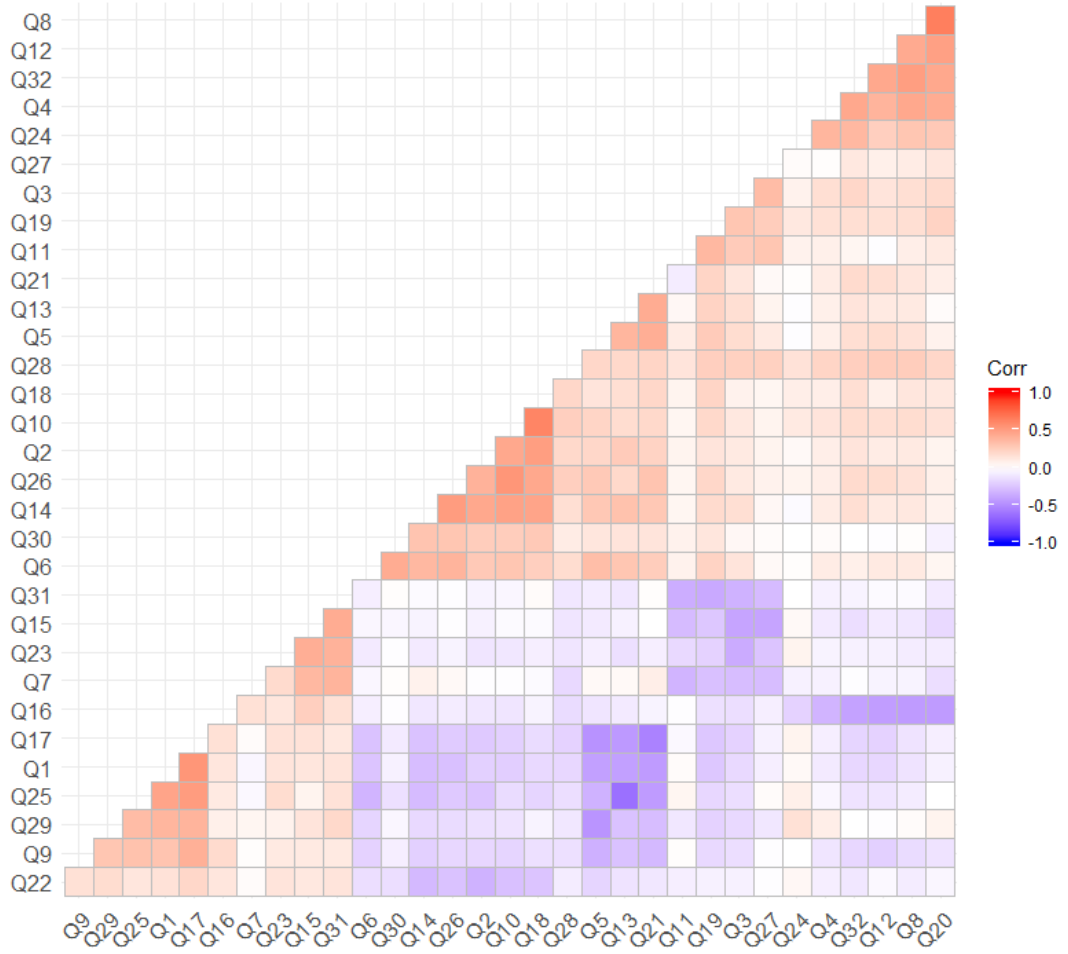
R Example

Libraries:

- FactoMineR
- factoextra
- psych

R Example

```
correlations <- cor(humor_df[, 1:32])
ggcorrplot(correlations, hc.order = TRUE, type = "lower")
```



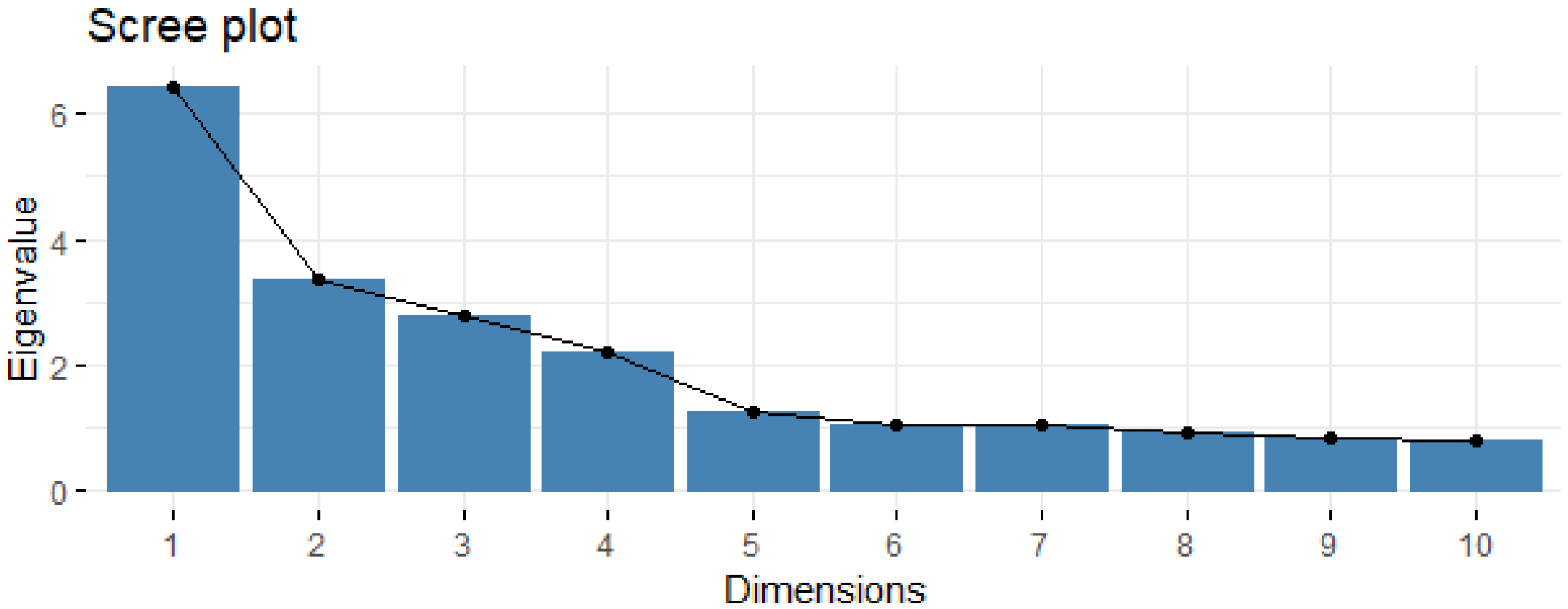
R Example

```
pca_solution <- PCA(humor_df[, 1:32],  graph = F)
pca_solution$eig[1:10,]
```

```
        eigenvalue percentage of variance cumulative percentage of variance
comp 1   6.4251214                20.078504                          20.07850
comp 2   3.3496777                10.467743                          30.54625
comp 3   2.7803371                 8.688553                          39.23480
comp 4   2.2007916                 6.877474                          46.11227
comp 5   1.2658822                 3.955882                          50.06816
comp 6   1.0430514                 3.259536                          53.32769
comp 7   1.0291327                 3.216040                          56.54373
comp 8   0.9298411                 2.905753                          59.44948
comp 9   0.8310916                 2.597161                          62.04665
comp 10  0.7873079                 2.460337                          64.50698
```

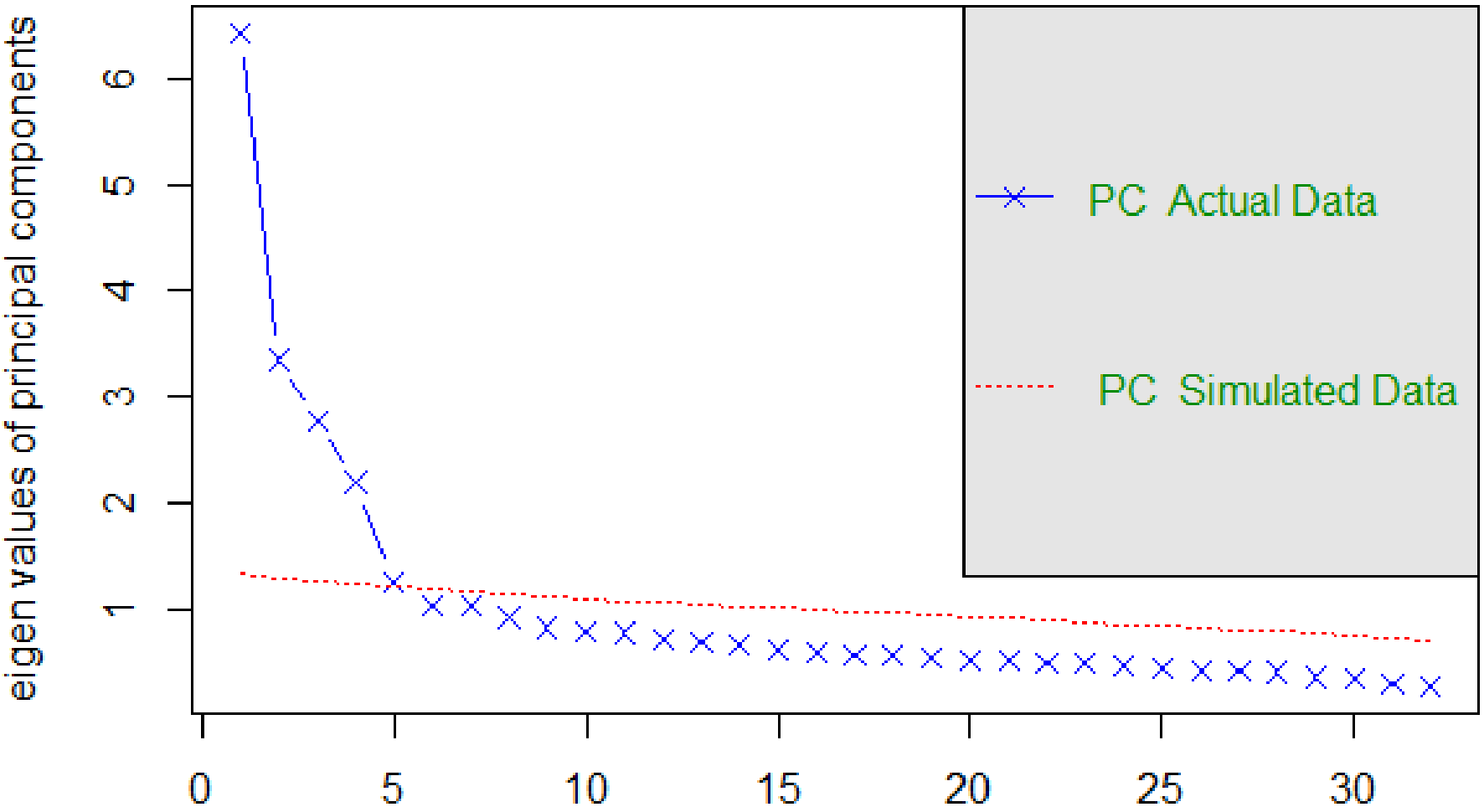
R Example

```
fviz_screeplot(pca_solution, choice="eigenvalue", ncp=10)
```



R Example

```
fa.parallel(correlations, n.obs = nrow(humor_df), fa='pc')
```



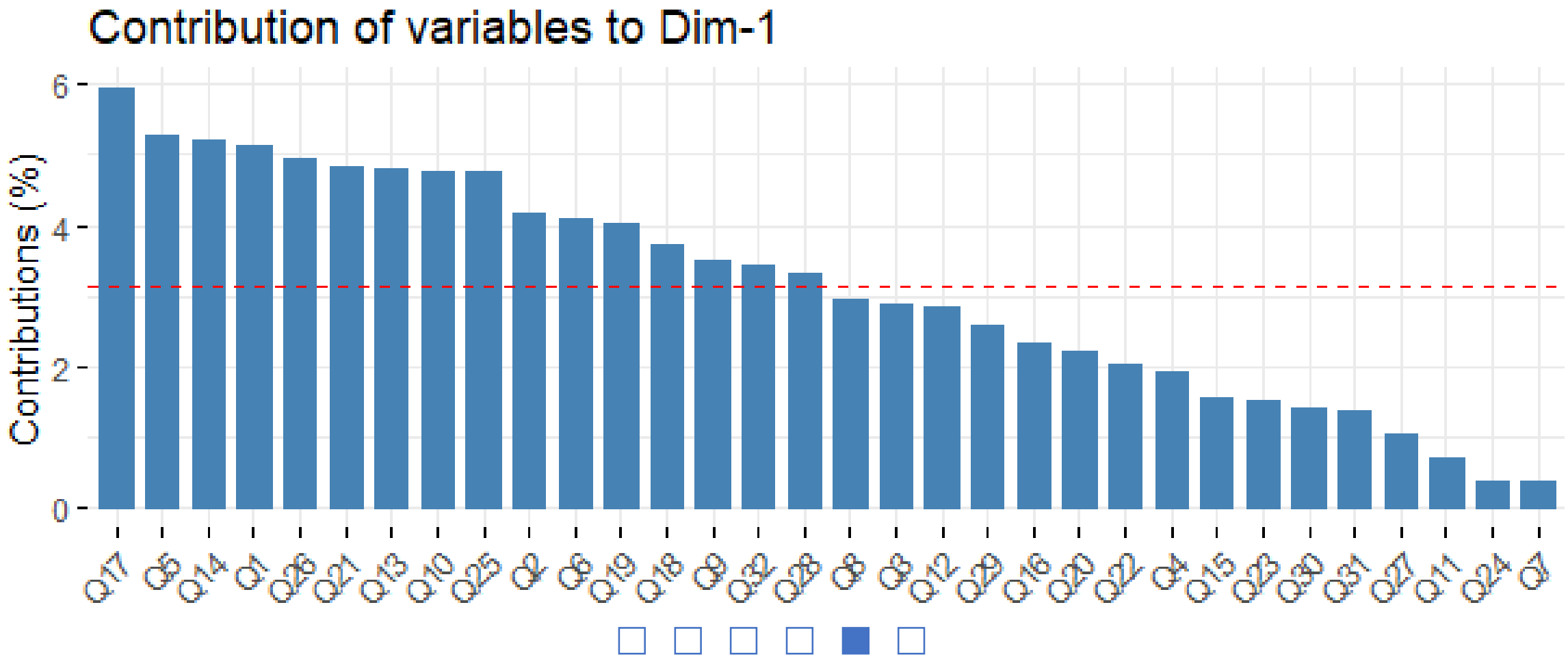
R Example

```
pca_solution$var$contrib
```

	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5
Q1	5.1194825	1.8982190	0.32582368	5.52898208	0.007058859
Q2	4.1747134	1.5868834	0.56863544	6.38054326	3.295721802
Q3	2.8728090	2.8309920	4.93548925	0.12309801	0.399124379
Q4	1.9268517	6.5242143	4.81237241	0.29001586	0.186699697
Q5	5.2697039	1.6617637	0.22039164	3.03789631	2.103176089
Q6	4.0815249	1.9283191	0.03215734	1.30397224	4.038561163
Q7	0.3840574	5.0973447	7.14419458	1.65984769	0.244074504
Q8	2.9531988	7.4096755	5.51440919	0.67385397	0.099283054
Q9	3.4948936	0.4236254	0.01749991	3.83240527	2.428959927
Q10	4.7655356	0.7912757	2.08228466	10.04069581	0.576040695
Q11	0.7050342	2.7987943	6.27562343	3.64147208	6.093660107
Q12	2.8240682	4.6837937	4.68779958	2.17690380	0.135934482
Q13	4.8133295	2.5289803	0.21393210	3.28366201	1.421871495
Q14	5.2083344	2.7775941	0.81185002	4.63999290	0.856600869
Q15	1.5438891	4.3426232	9.86039614	0.20381775	4.867182878
Q16	2.3266520	5.7205428	1.11250125	1.46154892	10.157047159
Q17	5.9327066	2.0024106	0.61368117	7.04118326	0.075293614
Q18	3.7442609	1.1394071	1.92973080	12.39274961	0.334107441
Q19	4.0237898	1.1895085	2.68186786	0.47072996	10.059822243
Q20	2.2040577	10.8107752	3.73920932	0.39511528	0.225015205
Q21	4.8232124	2.7880688	0.17224036	4.34297444	2.838159919
Q22	2.0146042	0.5847659	0.01093339	2.30041564	21.471364100
Q23	1.5216679	1.1744553	7.01899899	0.19113216	8.277947718
Q24	0.3902661	5.3457916	5.02464835	0.08105616	6.094411683
Q25	4.7643056	3.6644091	0.43642590	5.27054061	0.032014127
Q26	4.9573424	1.7821509	1.53212555	5.93076565	0.099902682
Q27	1.0446648	3.1785818	6.43295217	1.28628641	2.296042028
Q28	3.3324692	0.9368762	0.08119855	0.28600244	7.278929327
Q29	2.5860814	2.3603965	3.97019379	3.56271130	0.135115562
Q30	1.3932667	1.9235852	0.15256050	7.14249642	2.565807285
Q31	1.3835126	2.6681821	12.76043981	0.19394225	1.239516799
Q32	3.4197135	5.4459941	4.82743285	0.83319043	0.065553109

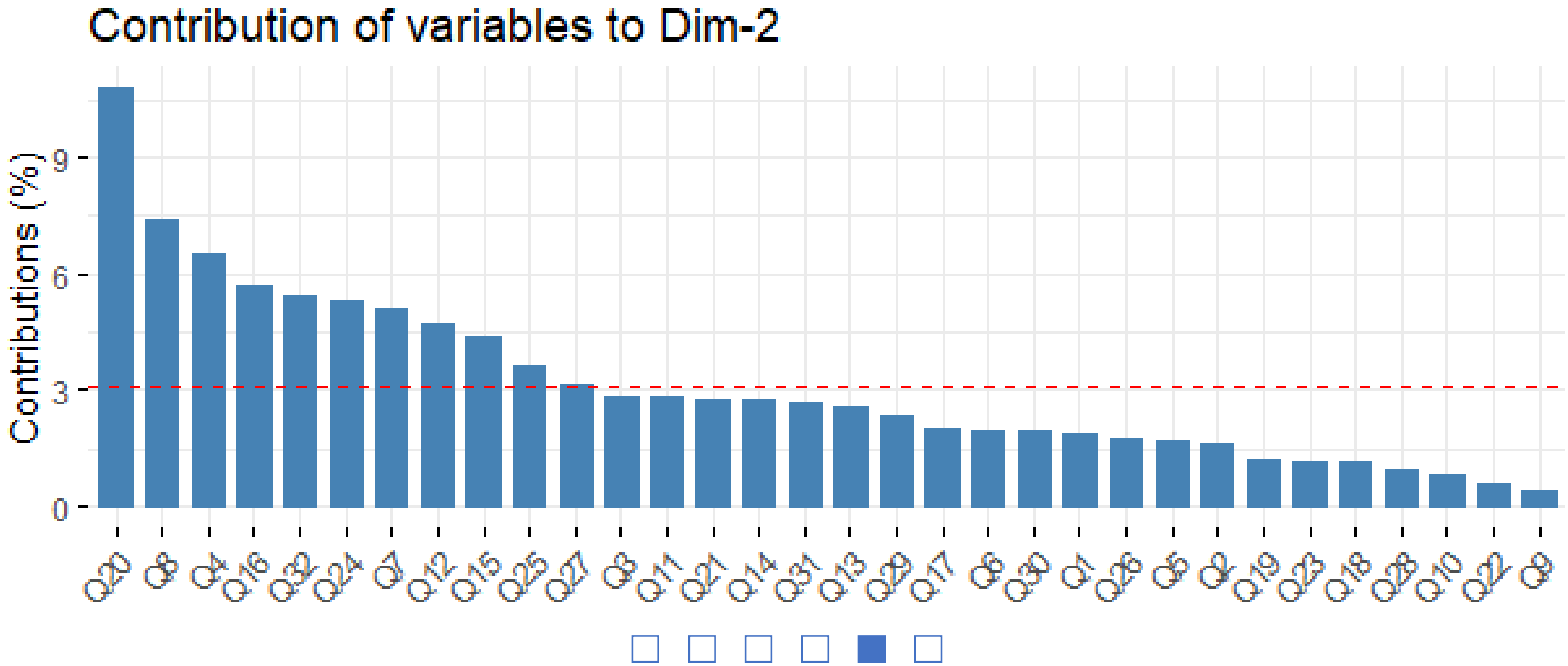
R Example

```
fviz_contrib(pca_solution, choice='var', axes = 1)
```



R Example

```
fviz_contrib(pca_solution, choice='var', axes = 2)
```



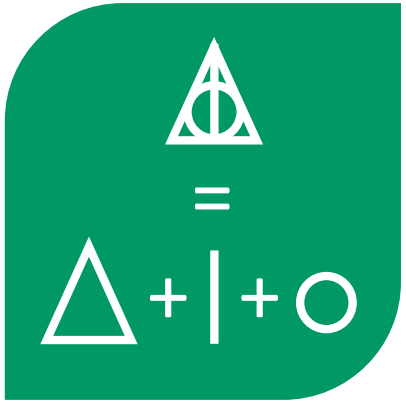
R Example

1. Affiliative Humor
2. Defeating Humor
3. Aggresive Humor
4. Enhancing Humor
5. Lack of Humor

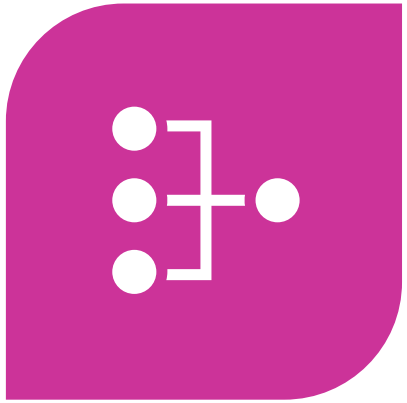
Other techniques

LDA

SVD



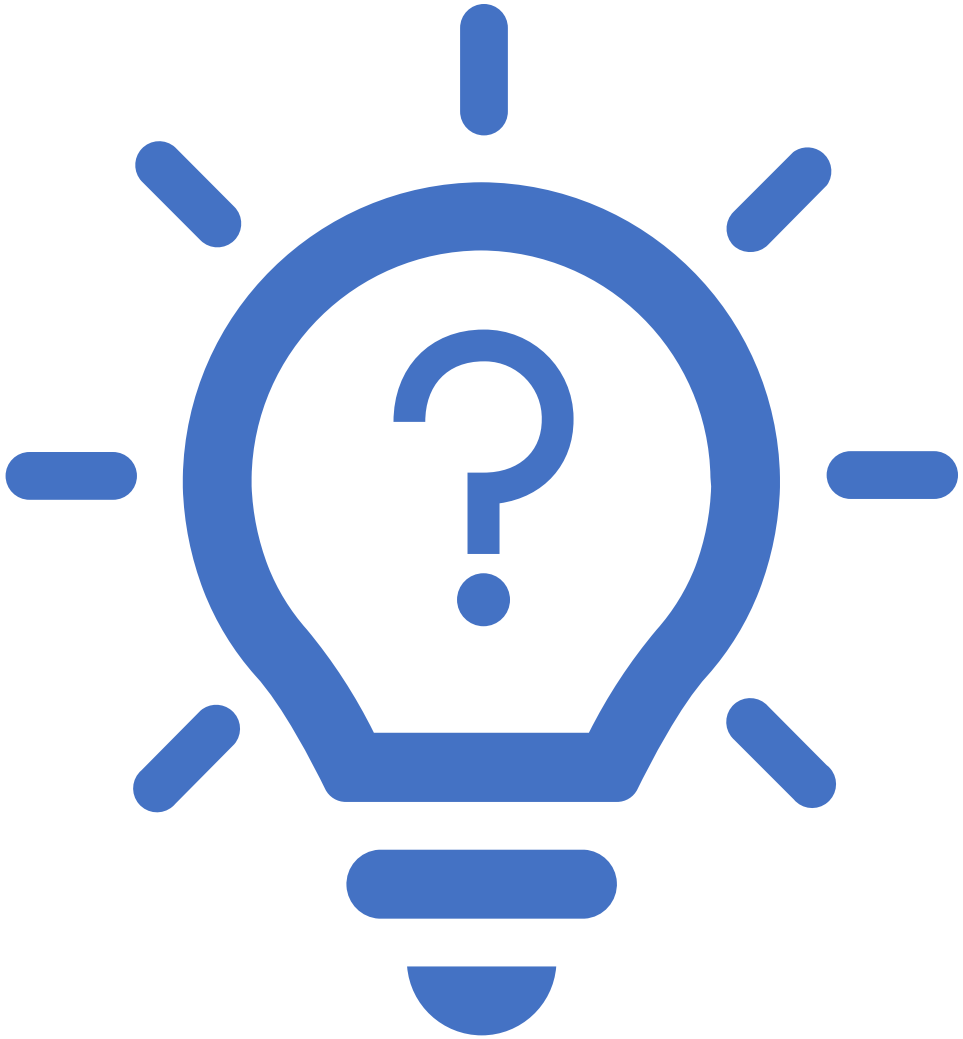
ICA



EFA

Thank you!

Questions?

ruben.guerrero.ramirez@gmail.com

linkedin.com/in/rubengura/