Portfolio 10, Study Group 10

Jesper Fischer, Kristian Severin, Lasse Hansen, Lærke Brædder, Sarah Hvid

5/4/2020

```
knitr::opts_chunk$set(fig.width=12, fig.height=8)
```

Loading packages/data

```
pacman::p_load(corpcor, GPArotation, psych, polycor, mvtnorm, pastecs, tidyverse,
fmsb, ggpubr)
```

```
df <- read.csv("Data/emp_all_all.csv")
```

Renaming column/row names

```
colnames(df) <- c(1:73)
rownames(df) <- c(1:104)
```

Testing for correlation

When dealing with factor analysis we need to have variables that correlate fairly well, but not perfectly. Therefore it is usefull to do a correlation matrix to check the pattern of relationships. If we have correlations that are greater than .9 we might face the problem of multicolinearity in the data. If these correlations are not existing >.3 we face a different problem, namely that a PCA might not be very usefull:

```
matrix_1 <- cor(df)
matrix_2 <- round(matrix_1, digits = 2)
```

To make this determination easier we will run a Barlett's test:

```
cortest.bartlett(matrix_2, n = 5329)
```

```
## $chisq
## [1] 355985.1
##
## $p.value
## [1] 0
##
## $df
## [1] 2628
```

For factor analysis to work we need some relationships between variables and if the R-matrix were an identity matrix then all correlation coefficients would be zero. Therefore, we want this test to be significant. A significant test tells us that the R-matrix is not an identity matrix; therefore, there are some relationships between the variables we hope to include in the analysis. For these data, Bartlett's test is highly significant, $\chi 2(2628) = 507902{,}8$, $p = 0$, and therefore factor analysis is appropriate.

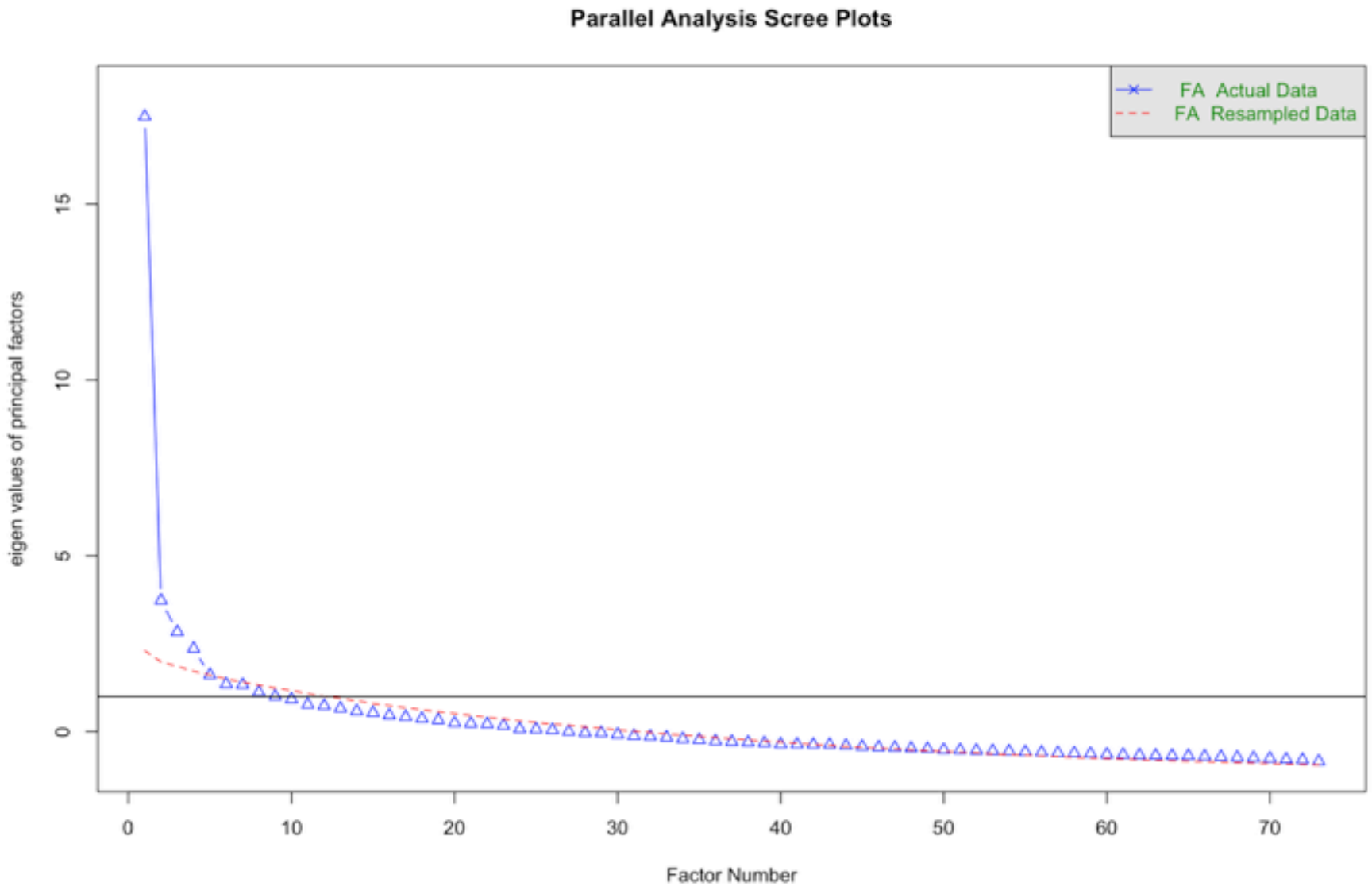
1) People at your agency disagree on how many interesting components/factors are present in the test, so they ask you, the factor analysis expert, to determine this. Please add your argument for the number you end on.

Horns mehtod:

To determine how we choose an interesting ammount of factors without anything not explaining so little variance that they are not relevant. We will do Horn's parallel analysis a statistical method used to determine the number of components to keep in a principal component analysis or factors to keep in an exploratory factor analysis:

```
fa.parallel(df, fm = "minires", fa = "fa", sim = F)
```

```
## factor method not specified correctly, minimum residual (unweighted least squar
es  used
```




```
## Parallel analysis suggests that the number of factors =  4  and the number of c
omponents =  NA
```

From the visualisation it is made clear that it is equally good to have either 4 or 5 components present in our analysis. At the 6th component it seems that there is not significantly more variance to explain in the data. Untill further analysis we will therefore run it with 4 components, which makes good explanatory sense, as Davis, M. (1980) explains that the test taps into four factors underlying empathy.

2) In order to short-list candidates for the position, your job is to find the highest and lowest scoring candidate on each factor.

We will start by running the analysis with 4 underlying factors:

```
f_a <- fa(r = df, nfactors = 4,rotate = "oblimin")
```

We have chosen a oblique rotation as we expect our components to correlate somewhat, as they are all explaining some sort of empathy.

After this we isolate the loading scores and put them into a dataframe to further investigate how the different candidates are scoring on our 4 empathy factors:

```
scores <- as.data.frame(f_a$scores)
scores$ID <- c(1:104)
```

Then to be able to plot the different scores too visually see which participants score high/low on the different factors we make 4 subsets:

```
d_MR1 <- subset(scores, select = c(1, 5))
d_MR2 <- subset(scores, select = c(3, 5))
d_MR3 <- subset(scores, select = c(4, 5))
d_MR4 <- subset(scores, select = c(2, 5))
```

Afterwards we make 4 different plots for the peoples loading scores to the different components. But first we need to name them something, to be able to determine what properties people seem to display from the questionaire.

```
print(f_a$loadings, cut = 0.3, order = T)
```

```
##
## Loadings:
##    MR1    MR4    MR2    MR3
## 1          0.334
## 2
## 3                 0.440
## 4   0.435  0.304
## 5   0.364
## 6          0.317
## 7                 0.330  0.447
## 8
## 9   0.338  0.319
## 10  0.438
## 11         0.314
## 12         0.623
## 13
## 14         0.342
## 15         0.442
## 16  0.558
```

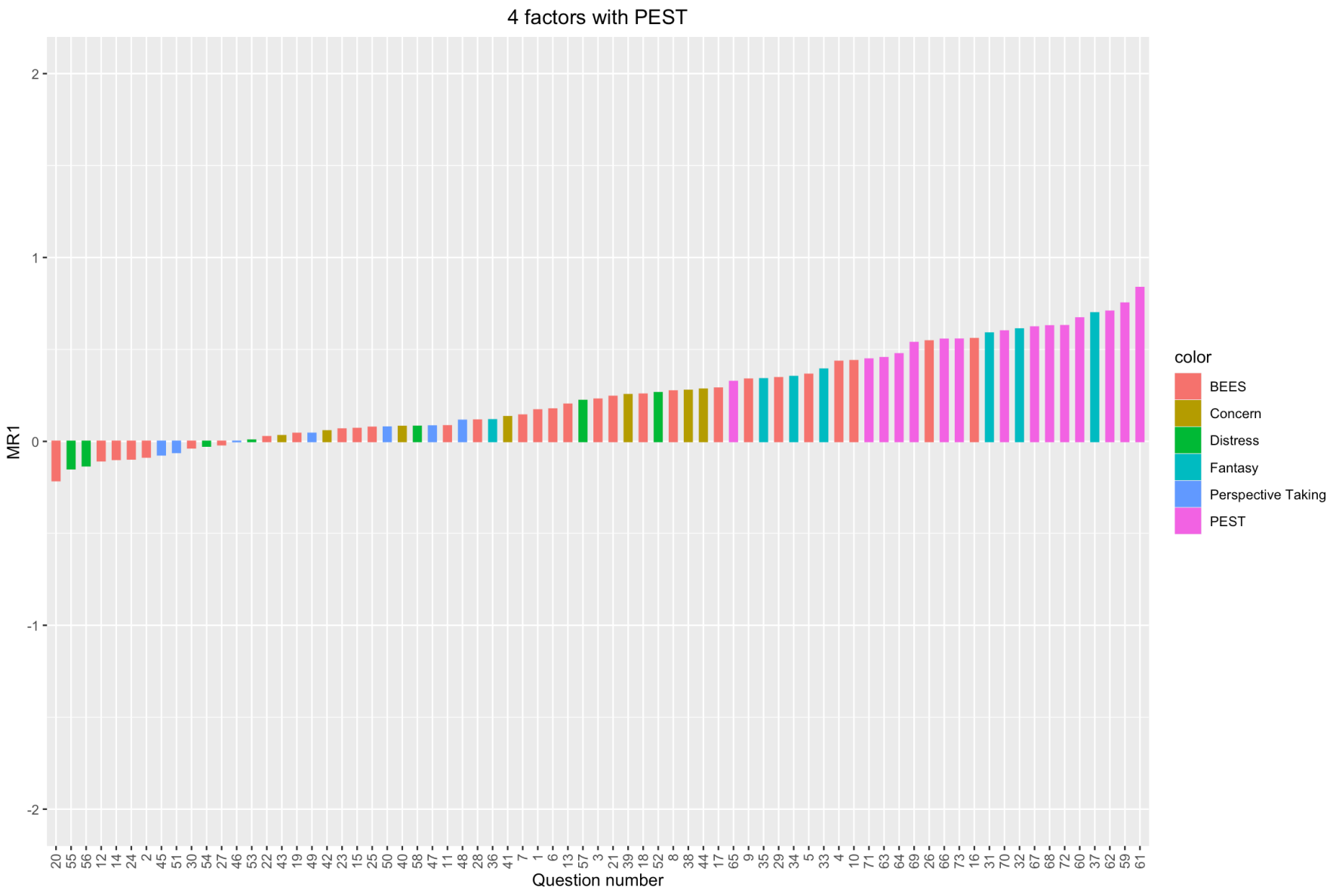
```
## 17                 0.469
## 18         0.513
## 19         0.550
## 20         0.500
## 21
## 22         0.569
## 23                 0.581
## 24         0.341        -0.327
## 25         0.455  0.456
## 26  0.546  0.376 -0.313
## 27         0.451
## 28
## 29  0.346          0.353
## 30         0.530
## 31  0.589
## 32  0.611
## 33  0.392
## 34  0.352
## 35  0.340          0.351
## 36         0.435
## 37  0.699
## 38                 0.466
## 39                 0.341
## 40                 0.307
## 41         0.500
## 42         0.465
## 43         0.574  0.328
## 44         0.557
## 45                 0.738
## 46                       -0.320
## 47                 0.529
## 48                 0.686
## 49                       -0.321
## 50                 0.423
## 51                 0.606
## 52                        0.367
## 53                        0.603
## 54                        0.840
## 55
## 56                        0.587
## 57                        0.324
## 58                        0.605
## 59  0.751
## 60  0.671
## 61  0.836
## 62  0.707
## 63  0.455
## 64  0.476
## 65  0.325
## 66  0.555
## 67  0.621
## 68  0.628
## 69  0.537
## 70  0.600
```

```
## 71  0.447
## 72  0.629
## 73  0.555
##
##                  MR1   MR4   MR2   MR3
## SS loadings    9.393 5.697 5.106 3.485
## Proportion Var 0.129 0.078 0.070 0.048
## Cumulative Var 0.129 0.207 0.277 0.324
```

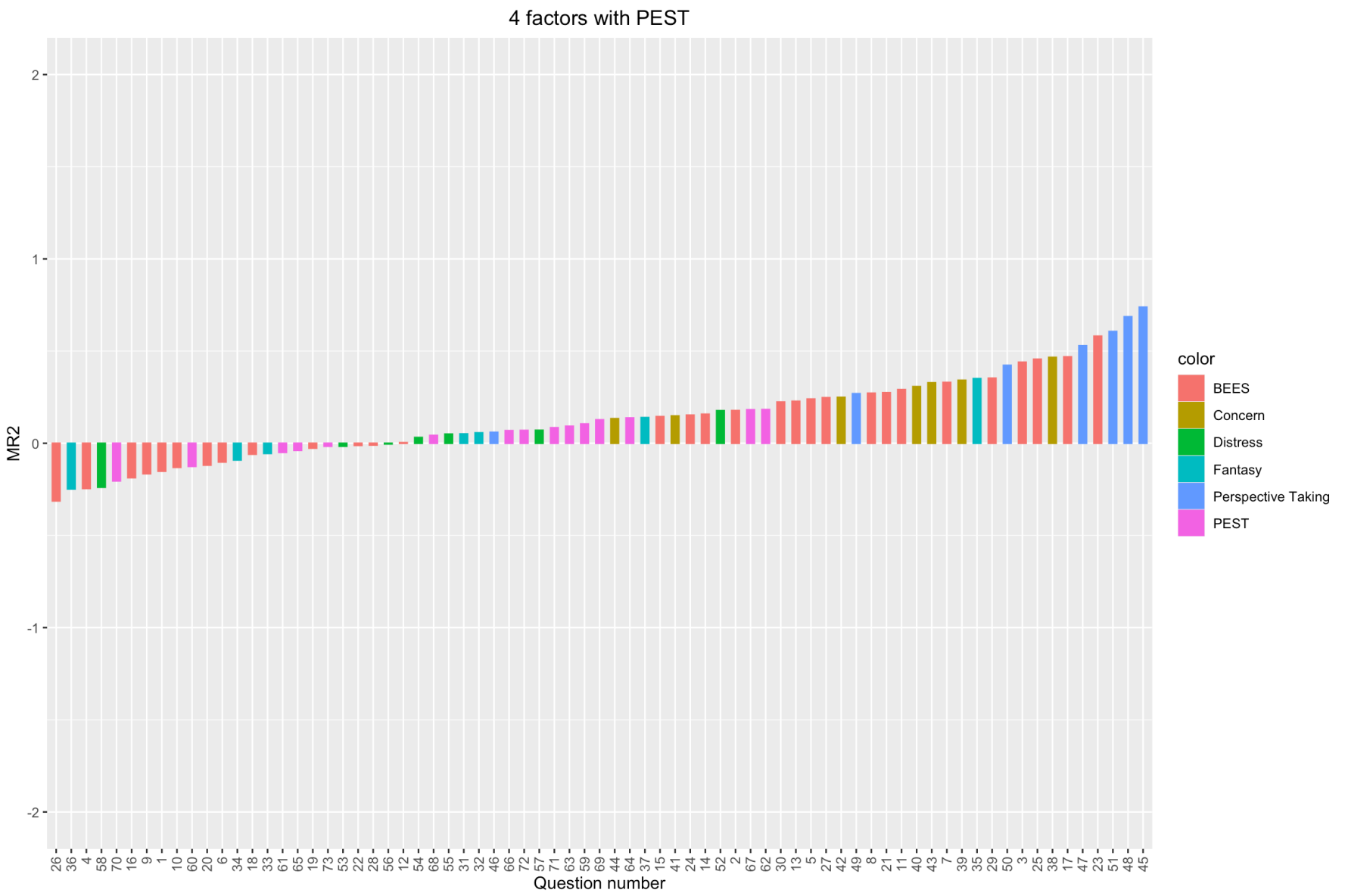
But this is a bit hard to comprehend, so

```
load <-  as.data.frame(f_a$loadings[,1:4]) %>%
            mutate(question_number = seq(73)) %>%
            mutate(color = ifelse(question_number >= 59, "PEST",
                      ifelse(question_number >= 52, "Distress",
                      ifelse(question_number >= 45, "Perspective Taking",
                      ifelse(question_number >= 38, "Concern",
                      ifelse(question_number >= 31, "Fantasy",
                      ifelse(question_number < 31, "BEES", ""))))))))
```

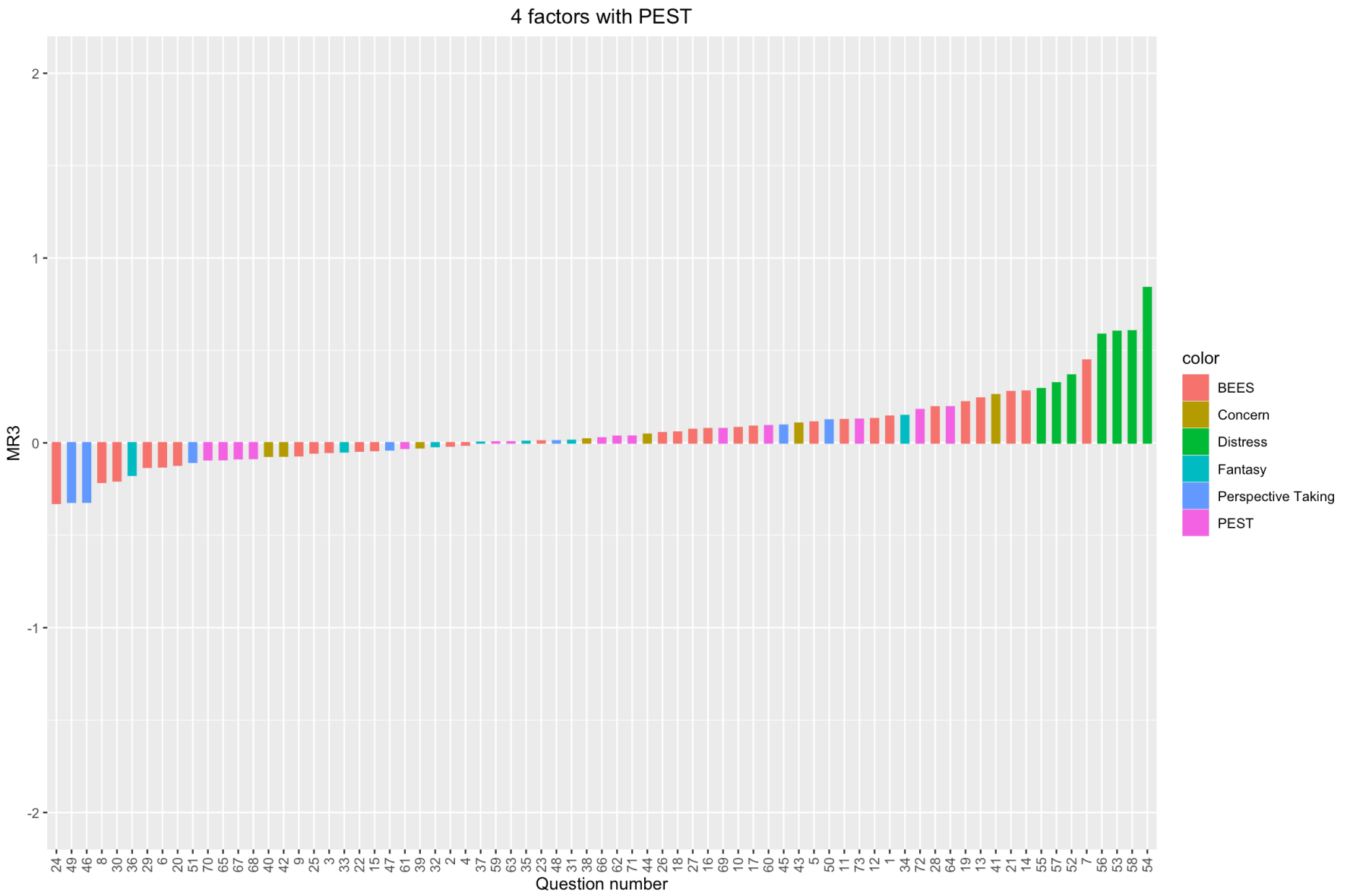
```
load %>% arrange(MR1) %>% mutate(question_number = factor(question_number, levels
= question_number)) %>%
            ggplot(., aes(question_number, MR1, fill = color, colour = color)) +
            geom_col(width = 0.5) + ylim(-2,2) + ggtitle("4 factors with PEST") +
xlab("Question number") +
            rotate_x_text() + theme(plot.title = element_text(hjust = 0.5))
```



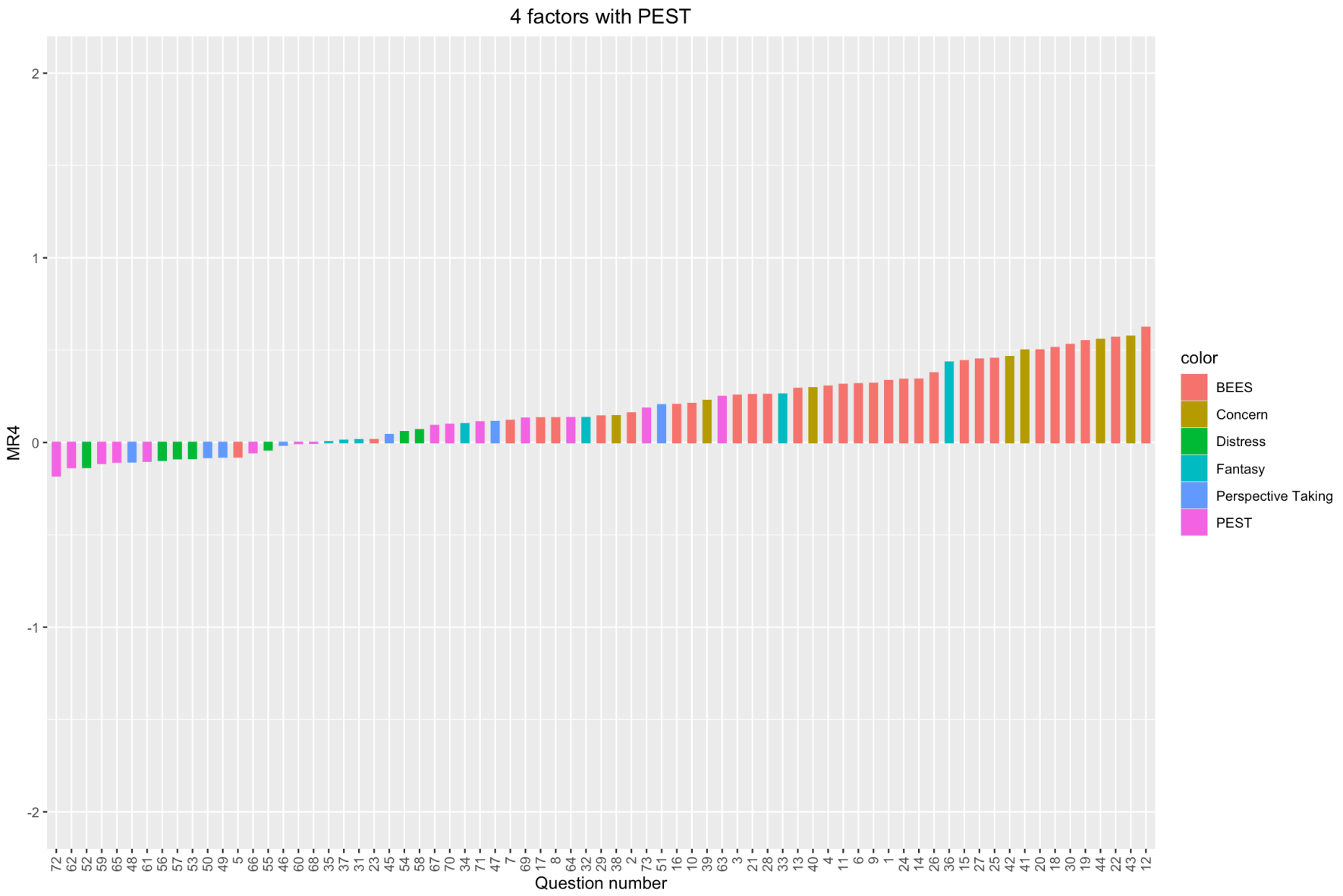
```
load %>% arrange(MR2) %>% mutate(question_number = factor(question_number, levels
= question_number)) %>%
            ggplot(., aes(question_number, MR2, fill = color, colour = color)) +
            geom_col(width = 0.5) + ylim(-2,2) + ggtitle("4 factors with PEST") +
xlab("Question number") +
            rotate_x_text() + theme(plot.title = element_text(hjust = 0.5))
```




```
load %>% arrange(MR3) %>% mutate(question_number = factor(question_number, levels
= question_number)) %>%
            ggplot(., aes(question_number, MR3, fill = color, colour = color)) +
            geom_col(width = 0.5) + ylim(-2,2) + ggtitle("4 factors with PEST") +
xlab("Question number") +
            rotate_x_text() + theme(plot.title = element_text(hjust = 0.5))
```




```
load %>% arrange(MR4) %>% mutate(question_number = factor(question_number, levels
= question_number)) %>%
            ggplot(., aes(question_number, MR4, fill = color, colour = color)) +
            geom_col(width = 0.5) + ylim(-2,2) + ggtitle("4 factors with PEST") +
xlab("Question number") +
            rotate_x_text() + theme(plot.title = element_text(hjust = 0.5))
```



The first factor (MR1) has high loadings for the “fantasy” and physical empathy questions. The other questions (the questions from the BEES test) that have high loadings in this factor fall under this category (i.e. “I am rarely moved to tears while reading a book or watching a movie” and other questions about books or movies).

The second factor (MR2) has high loadings for the “perspective taking” questions from the IRI test. When looking at the other questions that have high loadings in this factor they also have a tendency to fall under this category (i.e. “Another’s happiness can be very uplifting for me”).

The third factor (MR3) has high loadings for the “personal distress” questions from the IRI test, and the other questions that have high loadings in this factor involve questions like “It would be extremely painful for me to have to convey very bad news to another”.

The fourth factor (MR4) has high loadings for the “empathic concern” questions from the IRI test, and the other questions that have high loadings in this factor involve questions like “I not affected easily by the strong emotions of people around me”).

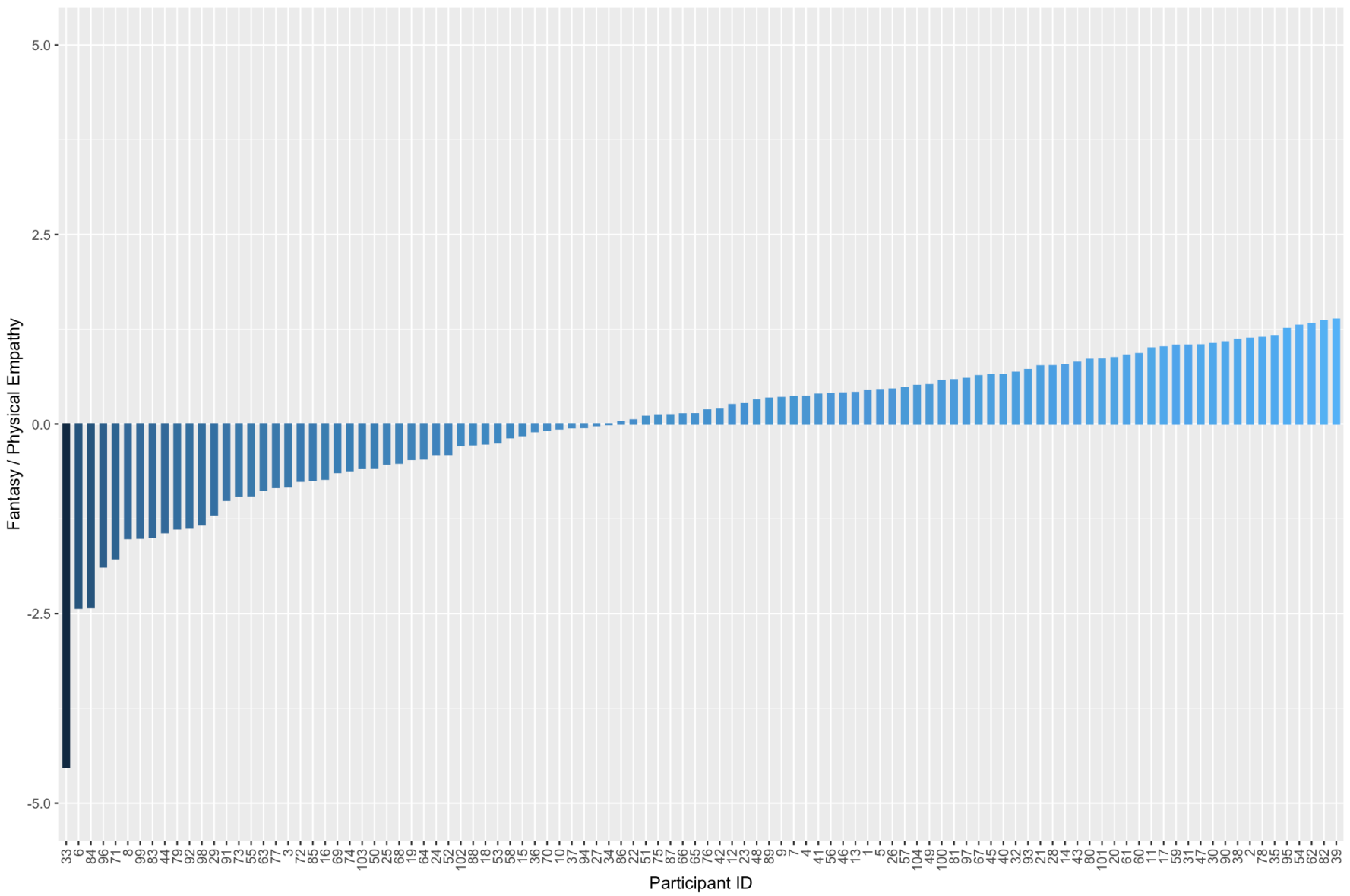
Hence, we have some different names for our different factors:

Fantasy/Physical empathy (MR1)

```
d_MR1 <- d_MR1 %>% arrange(MR1) %>% mutate(ID = factor(ID, levels = ID))

ggplot(d_MR1, aes(ID, MR1, fill = MR1, color = MR1)) +
    geom_col(stat = 'summary', fun.y = mean, width = 0.5) +
    ylim(-5,5) + xlab("Participant ID") + ylab("Fantasy / Physical Empathy") +
    rotate_x_text() + theme(legend.position = "none")
```

```
## Warning: Ignoring unknown parameters: stat, fun.y
```

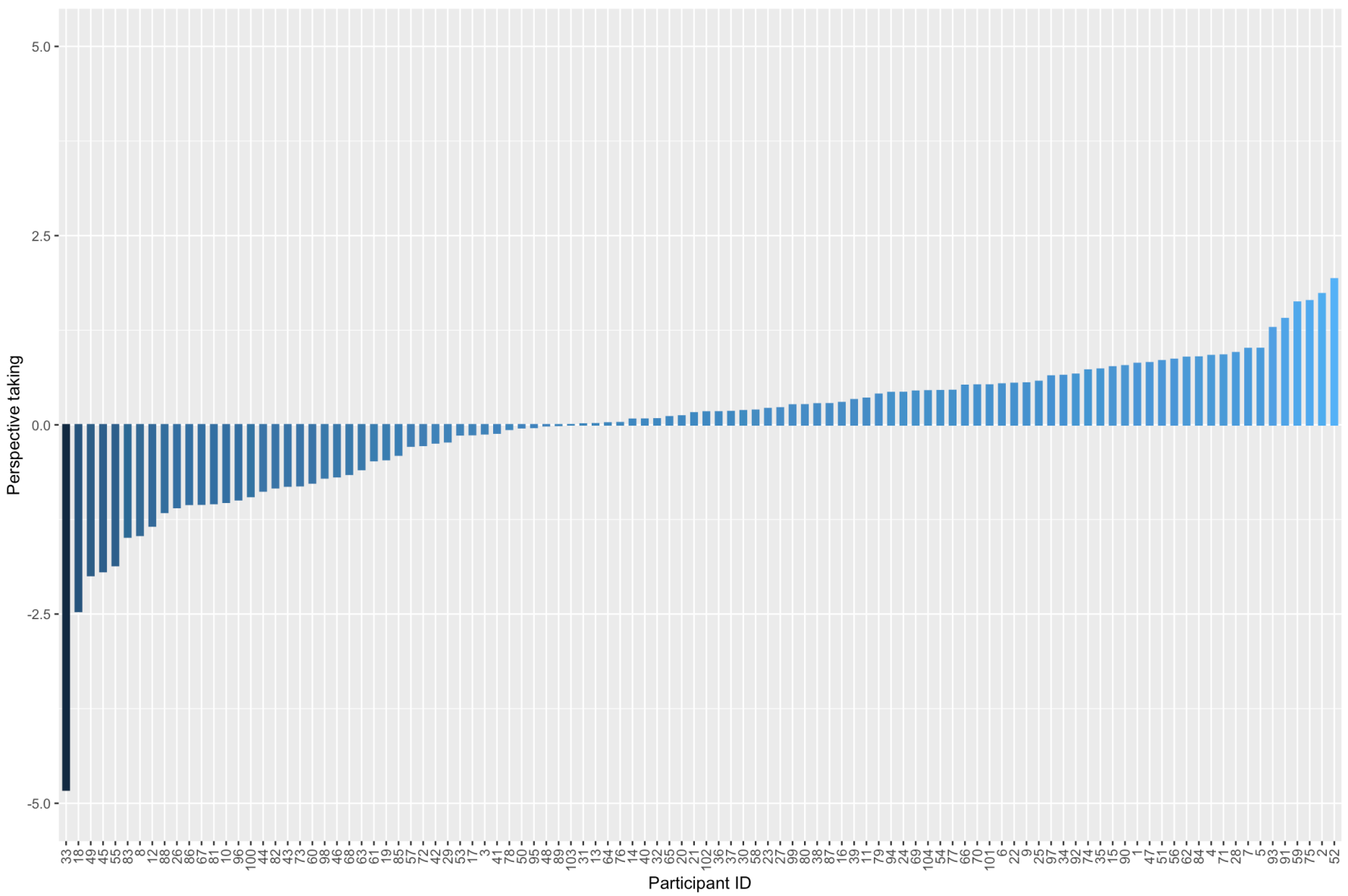


Perspective taking (MR2)

```
d_MR2 <- d_MR2 %>% arrange(MR2) %>% mutate(ID = factor(ID, levels = ID))

ggplot(d_MR2, aes(ID, MR2, fill = MR2, color = MR2)) +
    geom_col(stat = 'summary', fun.y = mean, width = 0.5) + ylim(-5,5) + xlab("Par
ticipant ID") + ylab("Perspective taking") +
    rotate_x_text() + theme(legend.position = "none")
```

```
## Warning: Ignoring unknown parameters: stat, fun.y
```

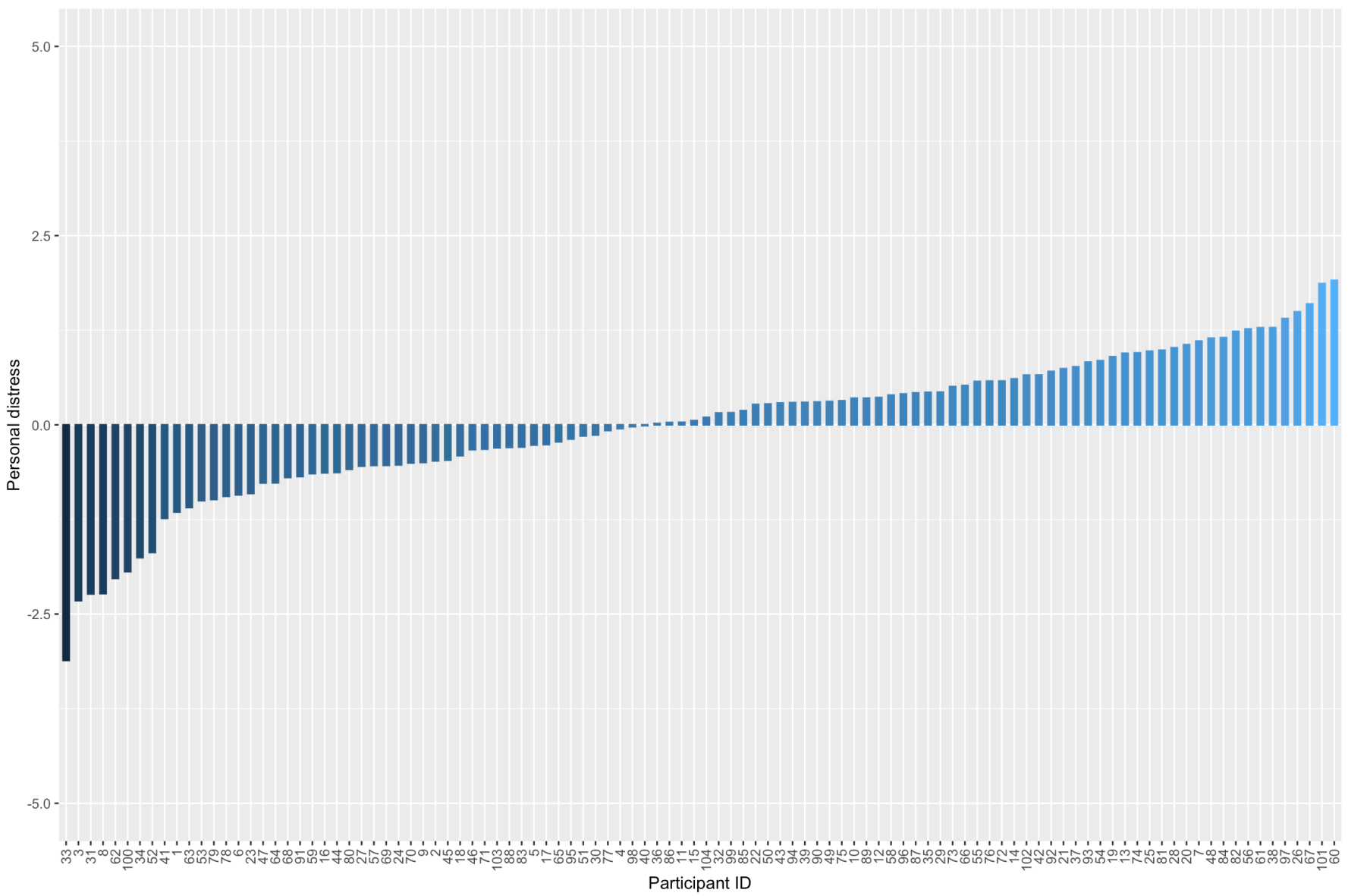


Personal distress (MR3)

```
d_MR3 <- d_MR3 %>% arrange(MR3) %>% mutate(ID = factor(ID, levels = ID))

ggplot(d_MR3, aes(ID, MR3, fill = MR3, color = MR3)) +
    geom_col(stat = 'summary', fun.y = mean, width = 0.5) + ylim(-5,5) +
    xlab("Participant ID") + ylab("Personal distress") + rotate_x_text() + theme(l
egend.position = "none")
```

```
## Warning: Ignoring unknown parameters: stat, fun.y
```

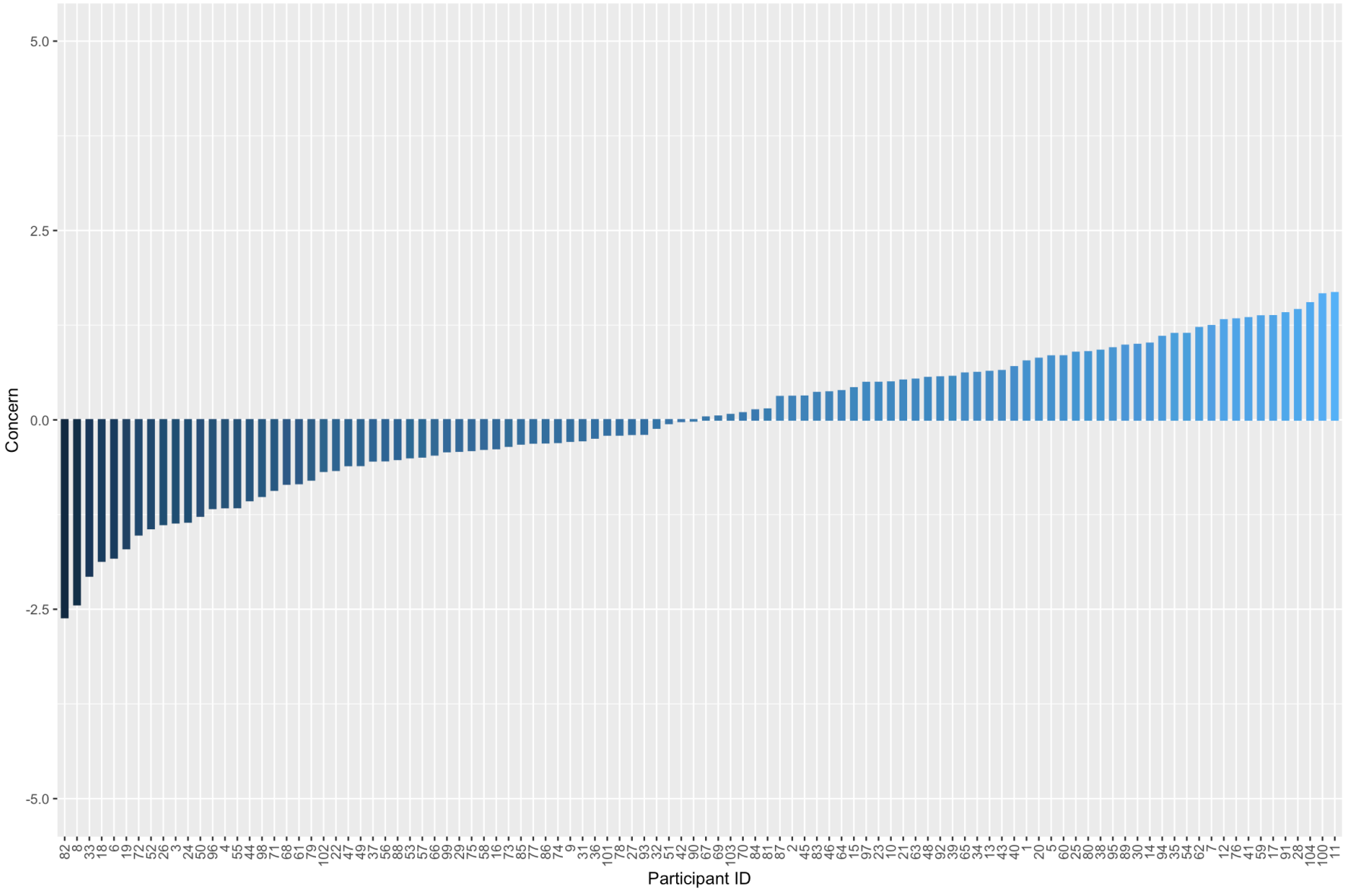


Concern (MR4)

```
d_MR4 <- d_MR4 %>% arrange(MR4) %>% mutate(ID = factor(ID, levels = ID))

ggplot(d_MR4, aes(ID, MR4, fill = MR4, color = MR4)) +
    geom_col(stat = 'summary', fun.y = mean, width = 0.5) + ylim(-5,5) +
    xlab("Participant ID") + ylab("Concern") +
    rotate_x_text() + theme(legend.position = "none")
```

```
## Warning: Ignoring unknown parameters: stat, fun.y
```



There might be a concern with participant 33 :).

On fantasy/physical empathy the highest scoring was candidate 39, and the lowest scoring was candidate 33. On perspective taking the highest scoring was candidate 52, and the lowest scoring was candidate 33. On personal distress the highest scoring was candidate 60, and the lowest scoring was candidate 33. On concern distress the highest scoring was candidate 11, and the lowest scoring was candidate 82.

3) Your boss asks you what you think of his new empathy test (The physical empathy test). Does it reallymeasure anything that the old scales cannot capture?

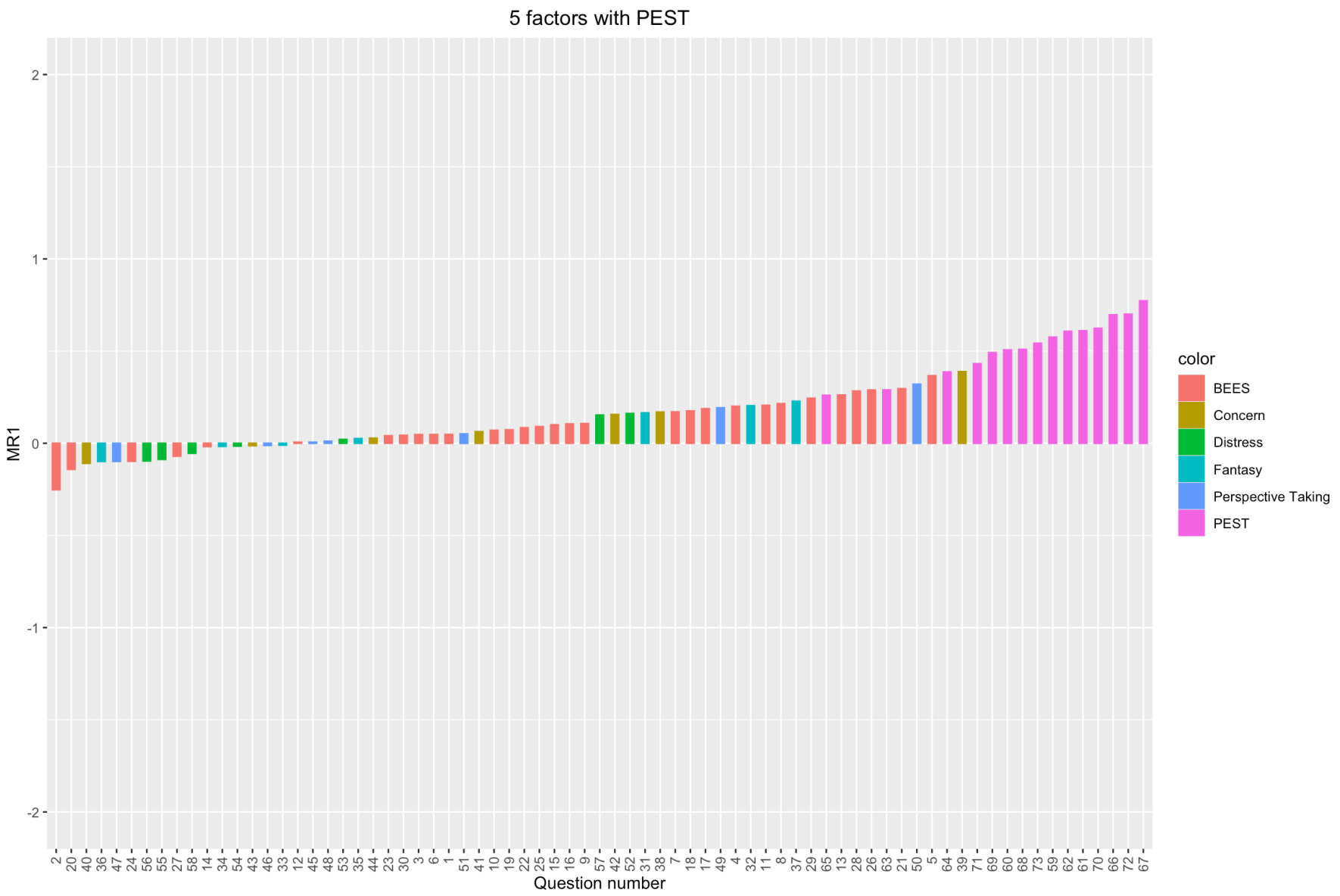
From Horn's parallel analysis we found that it is as plausible for there to be 5 underlying empathy factors (as the boss now proposes) as it is for there to be 4. (Look at question 1 for the scree plot). Therefore we will run the new analysis including the new physical empathy score with 5 factors instead of 4. And try to determine how this affects the questionaire and what is measures:

```
fa_new <- fa(r = df, nfactors = 5, rotate = "oblimin")
```

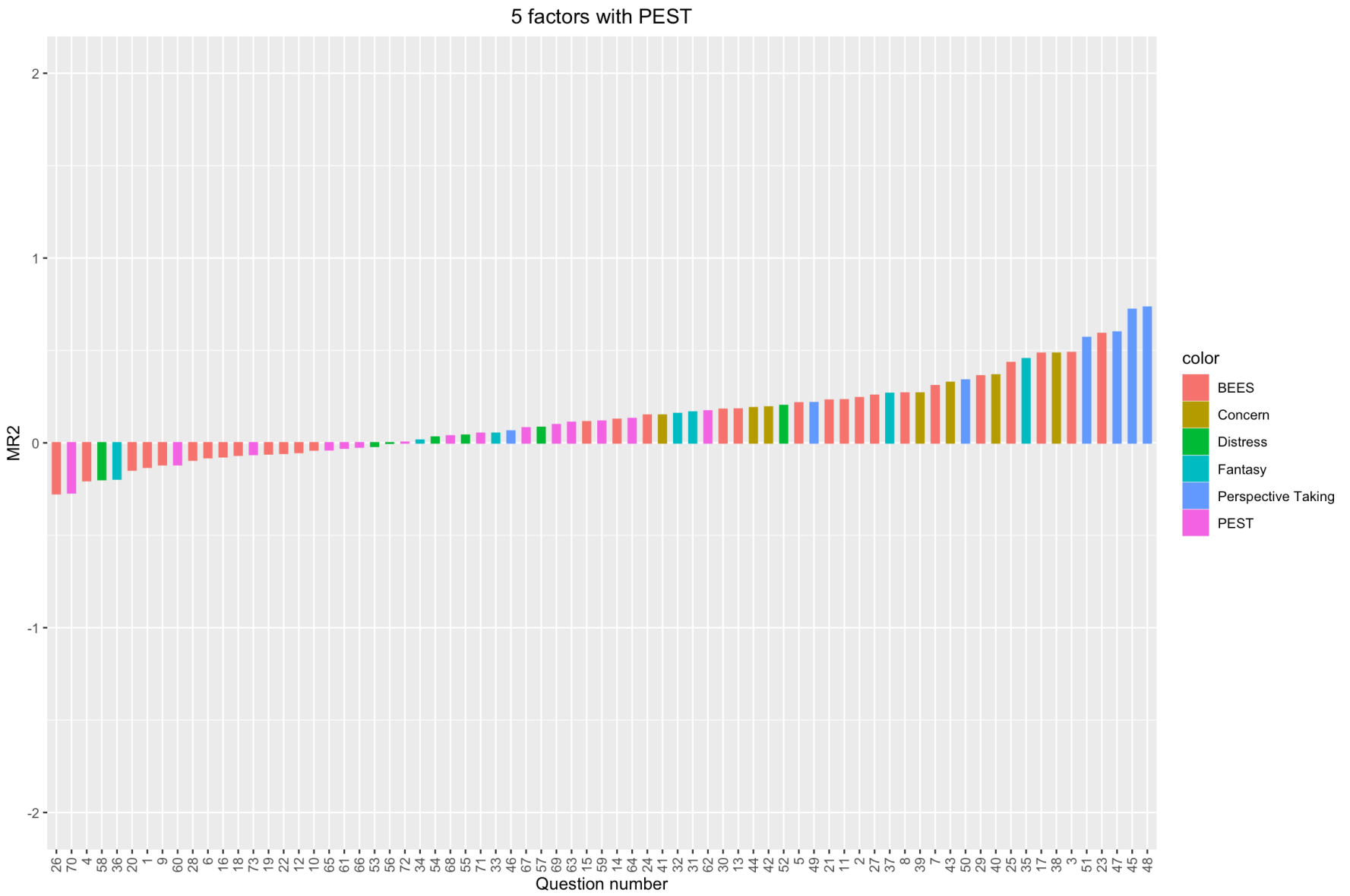
```
loading_new <-  as.data.frame(fa_new$loadings[,1:5]) %>%
            mutate(question_number = seq(73)) %>%
            mutate(color = ifelse(question_number >= 59, "PEST",
                      ifelse(question_number >= 52, "Distress",
                      ifelse(question_number >= 45, "Perspective Taking",
                      ifelse(question_number >= 38, "Concern",
                      ifelse(question_number >= 31, "Fantasy",
                      ifelse(question_number < 31, "BEES", ""))))))
```

We will then visualise each questionaire's loading on the different factors:

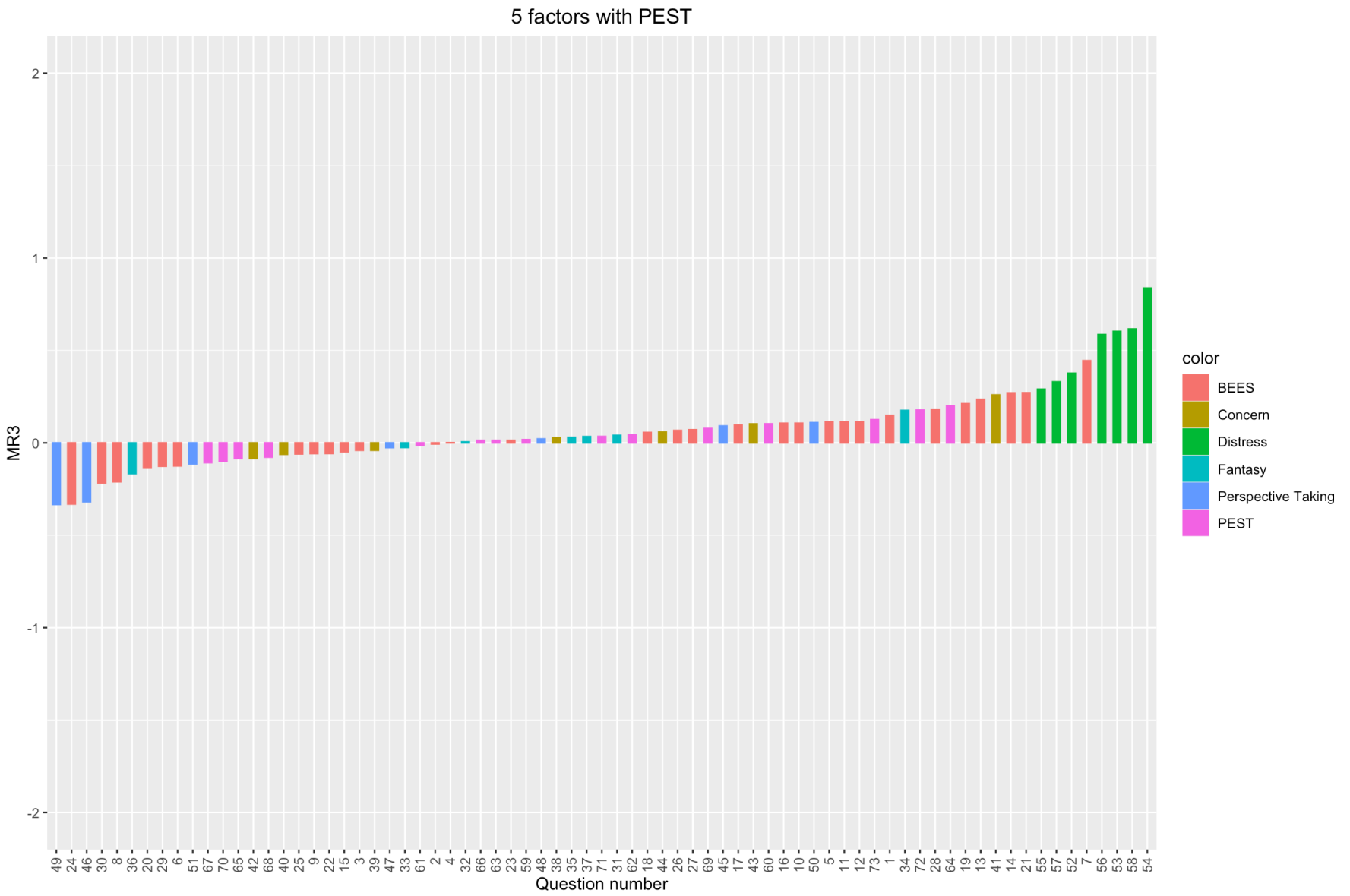
```
loading_new %>% arrange(MR1) %>% mutate(question_number = factor(question_number,
levels = question_number)) %>%
            ggplot(., aes(question_number, MR1, fill = color, colour = color)) +
            geom_col(width = 0.5) + ylim(-2,2) + ggtitle("5 factors with PEST") +
xlab("Question number") +
            rotate_x_text() + theme(plot.title = element_text(hjust = 0.5))
```



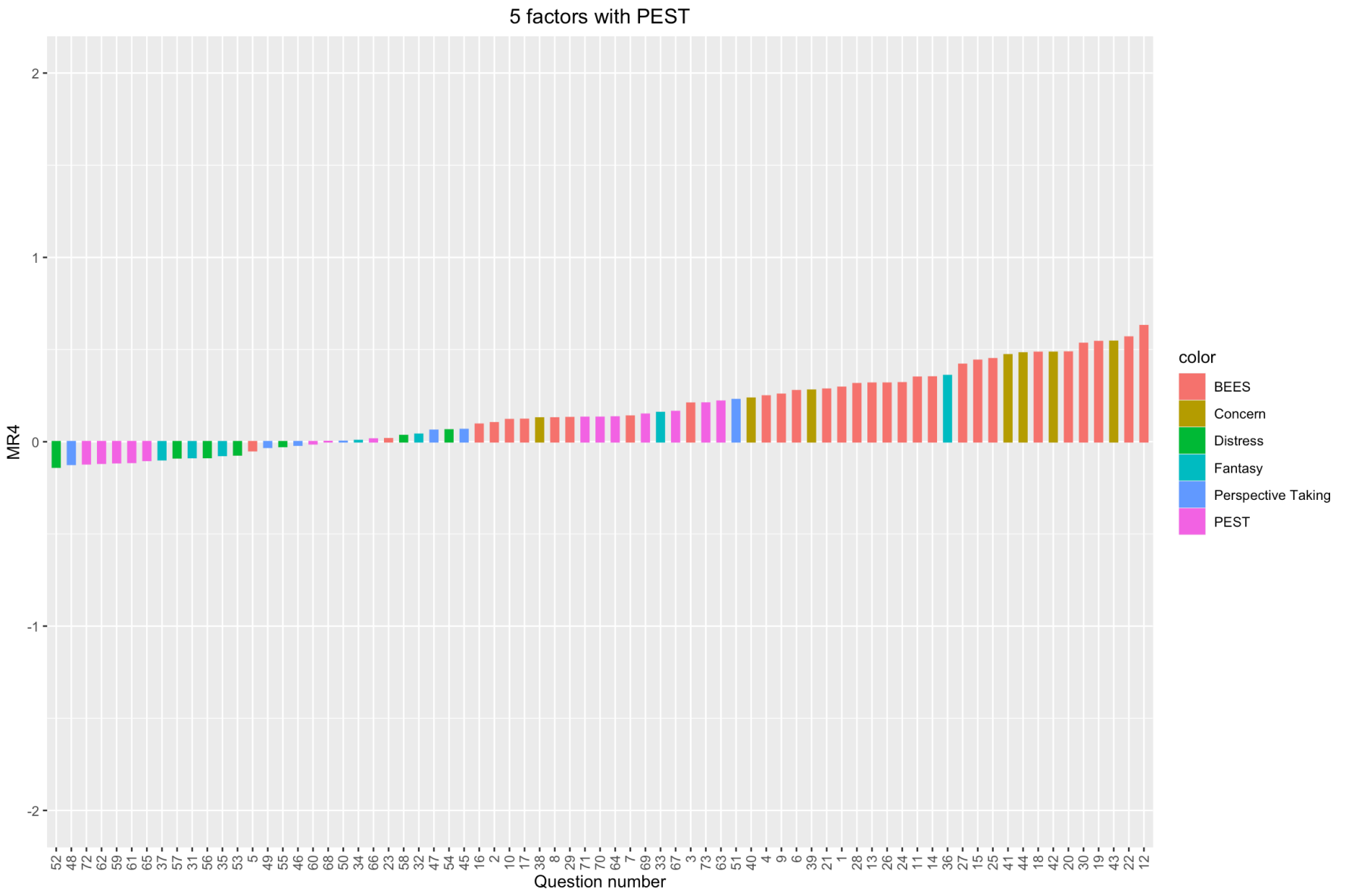

```
loading_new %>% arrange(MR2) %>% mutate(question_number = factor(question_number,
levels = question_number)) %>%
            ggplot(., aes(question_number, MR2, fill = color, colour = color)) +
            geom_col(width = 0.5) + ylim(-2,2) + ggtitle("5 factors with PEST") +
xlab("Question number") +
            rotate_x_text() + theme(plot.title = element_text(hjust = 0.5))
```



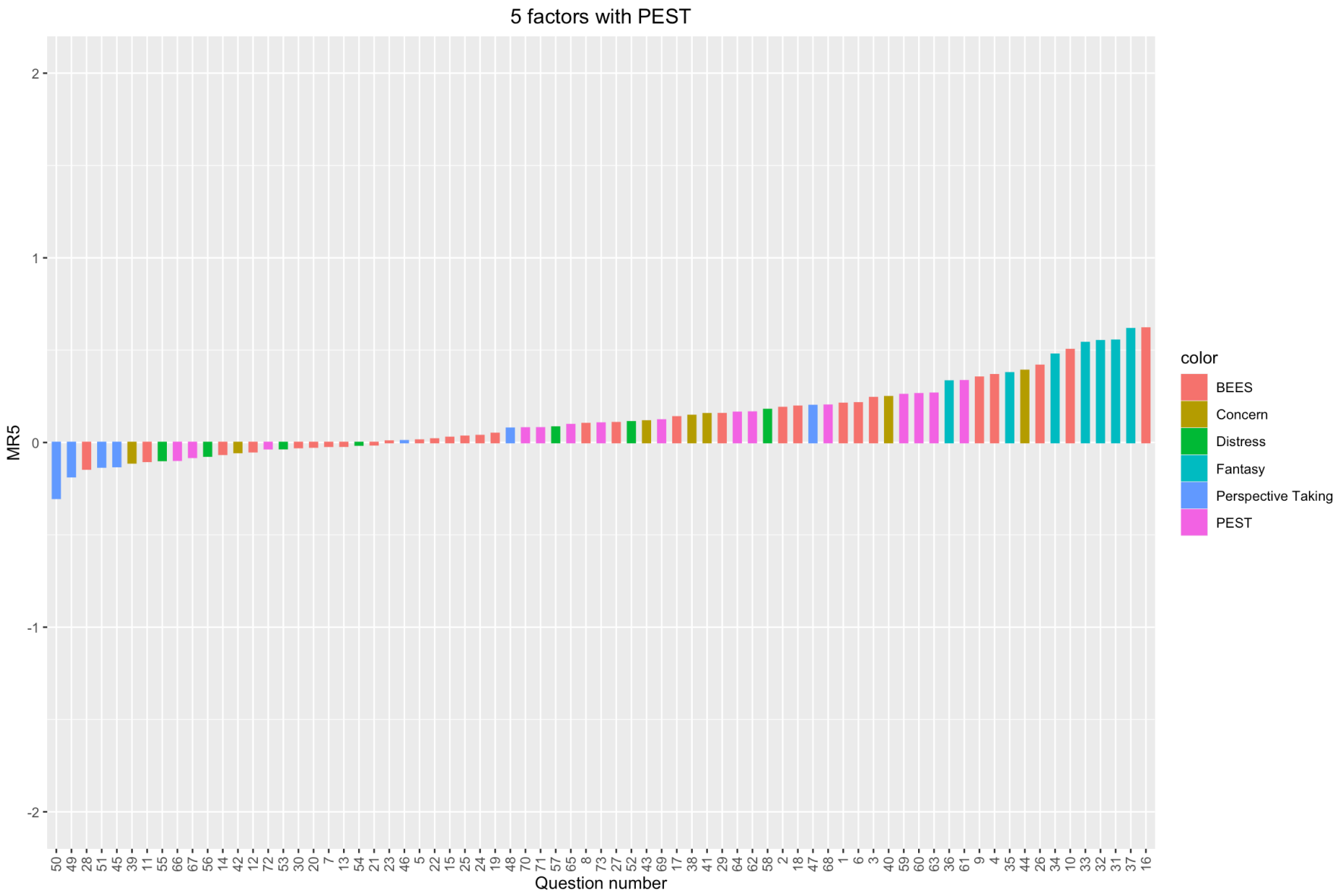

```
loading_new %>% arrange(MR3) %>% mutate(question_number = factor(question_number,
levels = question_number)) %>%
            ggplot(., aes(question_number, MR3, fill = color, colour = color)) +
            geom_col(width = 0.5) + ylim(-2,2) + ggtitle("5 factors with PEST") +
xlab("Question number") +
            rotate_x_text() + theme(plot.title = element_text(hjust = 0.5))
```




```
loading_new %>% arrange(MR4) %>% mutate(question_number = factor(question_number,
levels = question_number)) %>%
            ggplot(., aes(question_number, MR4, fill = color, colour = color)) +
            geom_col(width = 0.5) + ylim(-2,2) + ggtitle("5 factors with PEST") +
xlab("Question number") +
            rotate_x_text() + theme(plot.title = element_text(hjust = 0.5))
```




```
loading_new %>% arrange(MR5) %>% mutate(question_number = factor(question_number,
levels = question_number)) %>%
            ggplot(., aes(question_number, MR5, fill = color, colour = color)) +
            geom_col(width = 0.5) + ylim(-2,2) + ggtitle("5 factors with PEST") +
xlab("Question number") +
            rotate_x_text() + theme(plot.title = element_text(hjust = 0.5))
```



What we see here is that the different questions are distributed across the 5 factors almost the same as with 4 factors, except that now, the questions from the PEST test have highest loadings in a factor of their own instead of being combined with the fantasy category. We call this factor “physical empathy”, and judging by these results we conclude that the new test does measure something (that being physical empathy) that the old scale did not capture. So indeed you did a great job here boss.