1. The elbow point appears to be at 6, indicating that 6 factors are present
2. Using the Latent Root or Kaiser method, we drop any factor with an eigenvalue of less than 1, indicating that they explain the variance of less than one variable. That leaves us with 5 factors, which seems fairly reasonable.

```
3.

x18- Factor 1
x9- Factor 1
x16- Factor 1
x15- Factor 1, very low loading on all factors, no practical
significance
x6- Factor 2
x11- Factor 2, high crossloading with Factor 1
x13- Factor 2
x17- Factor 2, high crossloading with Factor 1
x12- Factor 3
x7- Factor 3
x10- Factor 3
x8- Factor 4
x14- Factor 4

x15 has no significant loadings, should be deleted (first edit). x15
and x17 both have very high crossloading, so they are also candidates
for deletion. I would delete x15 first, recalculate factors, and if
x11 and x17 are still crossloaded, remove x11, recalculate, and then
remove x17 if still crossloaded.
```

In [2]:
```
library(gdata)
```

```
gdata: read.xls support for 'XLS' (Excel 97-2004) files ENABLED.
```

In [19]:

```r
library(tidyverse)
```

```
── Attaching packages ─────────────────────────────────────────────────
─────────────────────────────── tidyverse 1.3.1 ──

✔ ggplot2 3.3.6     ✔ purrr   0.3.4
✔ tidyr   1.2.0     ✔ stringr 1.4.0
✔ readr   2.1.2     ✔ forcats 0.5.1

── Conflicts ──────────────────────────────────────────────────────────
────────────────────────── tidyverse_conflicts() ──
✖ dplyr::combine() masks gdata::combine()
✖ dplyr::filter()  masks stats::filter()
✖ dplyr::first()   masks gdata::first()
✖ purrr::keep()    masks gdata::keep()
✖ dplyr::lag()     masks stats::lag()
✖ dplyr::last()    masks gdata::last()
```

In [8]:

```r
hbat <- read.xls('HBAT(7).xls')
```

In [10]:

```r
head(hbat)
```

A data.frame: 6 × 24

	id	x1	x2	x3	x4	x5	x6	x7	x8	x9	…	x14	x15	x16
	<int>	<int>	<int>	<int>	<int>	<int>	<dbl>	<dbl>	<dbl>	<dbl>	…	<dbl>	<dbl>	<dbl>
1	1	2	0	1	1	1	8.5	3.9	2.5	5.9	…	4.7	4.3	5.0
2	2	3	1	0	0	0	8.2	2.7	5.1	7.2	…	5.5	4.0	3.9
3	3	3	0	1	1	1	9.2	3.4	5.6	5.6	…	6.2	4.6	5.4
4	4	1	1	1	1	0	6.4	3.3	7.0	3.7	…	7.0	3.6	4.3
5	5	2	0	1	0	1	9.0	3.4	5.2	4.6	…	6.1	4.5	4.5
6	6	1	1	0	1	0	6.5	2.8	3.1	4.1	…	5.1	9.5	3.6

In [15]:

```r
hbat %>% select(x6:x18) -> data
```

In [17]:

```r
data %>% select(!c(x15,x17)) -> data
```

In [18]:

```r
head(data)
```

A data.frame: 6 × 11

	x6	x7	x8	x9	x10	x11	x12	x13	x14	x16	x18

	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
1	8.5	3.9	2.5	5.9	4.8	4.9	6.0	6.8	4.7	5.0	3.7
2	8.2	2.7	5.1	7.2	3.4	7.9	3.1	5.3	5.5	3.9	4.9
3	9.2	3.4	5.6	5.6	5.4	7.4	5.8	4.5	6.2	5.4	4.5
4	6.4	3.3	7.0	3.7	4.7	4.7	4.5	8.8	7.0	4.3	3.0
5	9.0	3.4	5.2	4.6	2.2	6.0	4.5	6.8	6.1	4.5	3.5
6	6.5	2.8	3.1	4.1	4.0	4.3	3.7	8.5	5.1	3.6	3.3

In [27]:

```
cov(data) -> data.cov
```

In [28]:

```
eigen(data.cov) -> data.pc
data.pc
```

```
eigen() decomposition
$values
 [1] 4.80197660 3.09203585 2.48814620 1.47904219 1.18584833 0.6161498
```

We have five eigenvalues over 1, so by the Kaiser rule, we have 5 factors.

```
In [30]: prcomp(data, rank = 5, center = TRUE, scale = TRUE, retx = TRUE)
```

```
Standard deviations (1, .., p=11):
 [1] 1.8512081 1.5971527 1.3003755 1.0423800 0.7806562 0.7428888 0.63
36546
 [8] 0.4969422 0.4511688 0.3644744 0.3137308

Rotation (n x k) = (11 x 5):
           PC1         PC2         PC3        PC4         PC5
x6  -0.1337896 -0.31349802  0.06227164  0.6431362 -0.23166620
x7  -0.1659528  0.44650918 -0.23524791  0.2723803 -0.42228844
x8  -0.1576926 -0.23096734 -0.61095105 -0.1933931  0.02395667
x9  -0.4706836  0.01944394  0.21035078 -0.2063204 -0.02865743
x10 -0.1837350  0.36366471 -0.08809705  0.3178945  0.80387024
x11 -0.3867652 -0.28478056  0.11627864  0.2029023 -0.11667416
x12 -0.2036696  0.47069599 -0.24134210  0.2221772 -0.20437283
x13  0.1516886  0.41345650  0.05304529 -0.3335435 -0.24892601
x14 -0.2129336 -0.19167191 -0.59856398 -0.1853020  0.03292706
x16 -0.4372177  0.02639905  0.16892981 -0.2368536 -0.02675377
x18 -0.4730891  0.07305172  0.23262477 -0.1973299  0.03543294
```

x6- 4

x7- 2

x8- 3

x9- 1

x10- 5

x11- 1

x12- 2

x13- 2

x14- 3

x16- 1

x18- 1

Main problem: very few of the factors are significant. This indicates that we ought to drop the variables with least significant loading (like x11 and recalculate) and also drop those with problematic crossloading.