

HWZ CAS BigData - Cluster Analysis

thomas.stump@bluewin.ch

12/13/2018

CLUSTER ANALYTICS

thomas.stump@bluewin.ch HWZ Big Data Analysis

The following script depicts the cluster analysis used on random generated data, covering income vs. educational level (years in school)

Generate data

```
inc1 <- rnorm(100, mean = 10000, sd = 5000)
inc2 <- rnorm(100, mean = 40000, sd = 5000)
inc3 <- rnorm(100, mean = 70000, sd = 5000)
inc <- c(inc1, inc2, inc3)

head(inc)

## [1] 15717.762 7656.839 9407.780 9313.308 14222.383 14310.112

tail(inc)

## [1] 72217.52 72655.03 70289.24 79414.29 71742.88 60999.10

summary(inc)

##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    -1980   13425   40592   40246   67066   81806
```

Including Plots

You can also embed plots, for example:



Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.