306 First Draft: Sam, Jangmin, Matt

2025-04-01

```r
library(ggplot2)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```r
library(broom)
```

```r
bank_full <- read.csv("bank-full.csv", sep=";", header=TRUE)
bank_full <- bank_full |>
  select(age, job, marital, education, default, balance, housing, loan)
```

```r
bank_full$job <- as.character(as.numeric(factor(bank_full$job)))
bank_full$marital <- as.character(as.numeric(factor(bank_full$marital)))
bank_full$education <- as.character(as.numeric(factor(bank_full$education)))
bank_full$default <- as.character(as.numeric(factor(bank_full$default)))
bank_full$housing <- as.character(as.numeric(factor(bank_full$housing)))
bank_full$loan <- as.character(as.numeric(factor(bank_full$loan)))
```

```r
bank_full_management <- bank_full |>
  filter(job == 5)

set.seed(1)
random_sample <- bank_full[sample(nrow(bank_full), 100), ]

km_out = kmeans(bank_full, 2, nstart = 20)
km_clusters = km_out$cluster
df = table(km_clusters, bank_full$loan)




assignments <- augment(km_out, bank_full)
ggplot(data = assignments, aes(x = balance, y = age, color = .cluster)) +
  geom_point() + labs(color = "Cluster Assignment",
                      title = "K-Means Clustering Results with K = 2",
                      shape = "loan")
```

K–Means Clustering Results with K = 2