analysis_clustering

2025-05-15

```
library(tidyverse)
```

```
## Warning: Paket 'tidyverse' wurde unter R Version 4.4.2 erstellt

## Warning: Paket 'ggplot2' wurde unter R Version 4.4.3 erstellt

## Warning: Paket 'tibble' wurde unter R Version 4.4.2 erstellt

## Warning: Paket 'tidyr' wurde unter R Version 4.4.2 erstellt

## Warning: Paket 'readr' wurde unter R Version 4.4.2 erstellt

## Warning: Paket 'purrr' wurde unter R Version 4.4.3 erstellt

## Warning: Paket 'dplyr' wurde unter R Version 4.4.2 erstellt

## Warning: Paket 'stringr' wurde unter R Version 4.4.2 erstellt

## Warning: Paket 'forcats' wurde unter R Version 4.4.2 erstellt

## Warning: Paket 'lubridate' wurde unter R Version 4.4.2 erstellt

## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.5
## v forcats   1.0.0     v stringr   1.5.1
## v ggplot2   3.5.2     v tibble    3.2.1
## v lubridate 1.9.4     v tidyr     1.3.1
## v purrr     1.0.4
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(ggplot2)
library(cluster)
library(factoextra)
```

```
## Warning: Paket 'factoextra' wurde unter R Version 4.4.3 erstellt

## Welcome! Want to learn more? See two factoextra-related books at https://goo.gl/ve3WBa
```

```r
zaehlstellen <- read_csv("data/dauerzaehlstellen_location.csv")
```

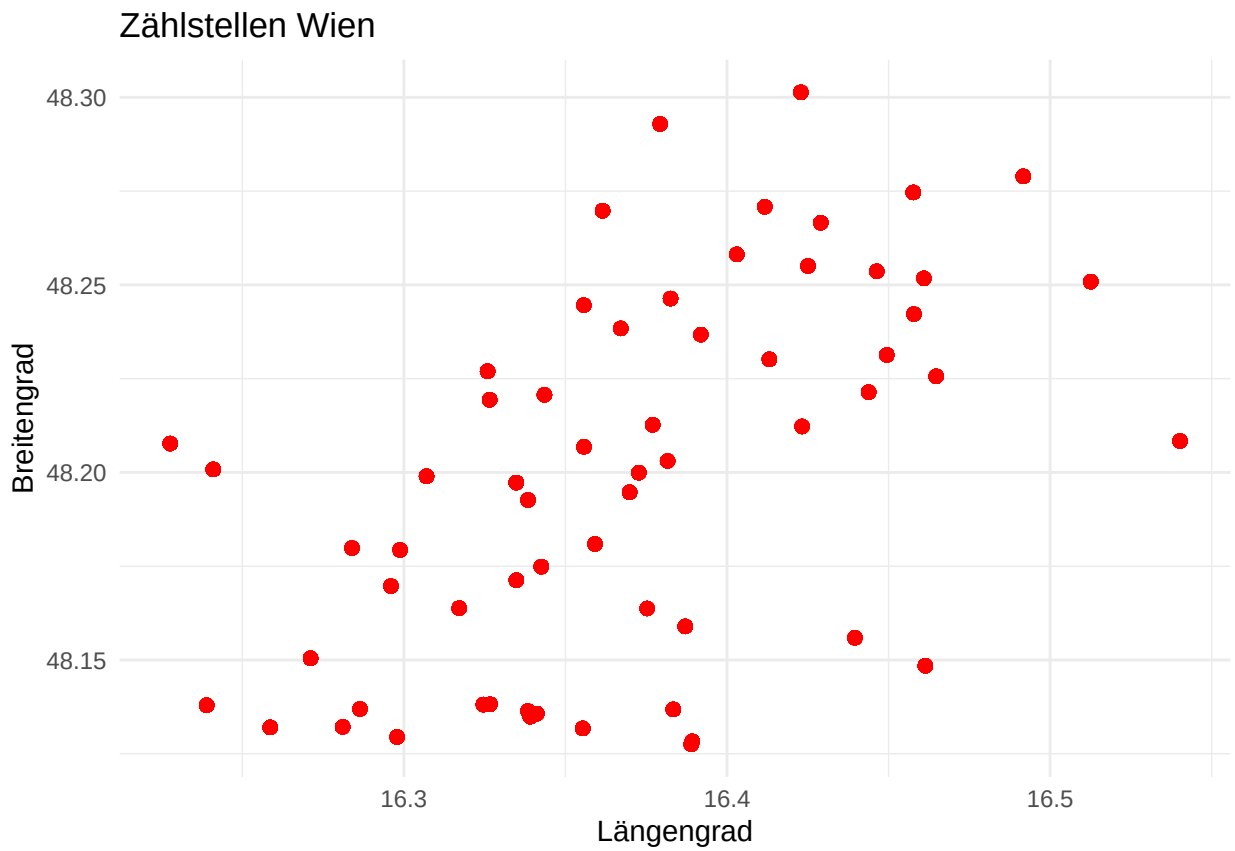
```
## Rows: 87 Columns: 11
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (5): ZNAME, STRNR, RICHTUNG_1, RICHTUNG_2, BEZIRK_NAME
## dbl (6): ZNR, LONGITUDE, LATITUDE, BEZIRK_PLZ, BEZIRK_NR, BEZIRK_CODE
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```r
verkehr <- read_csv("data/dauerzaehlstellen_data.csv")
```

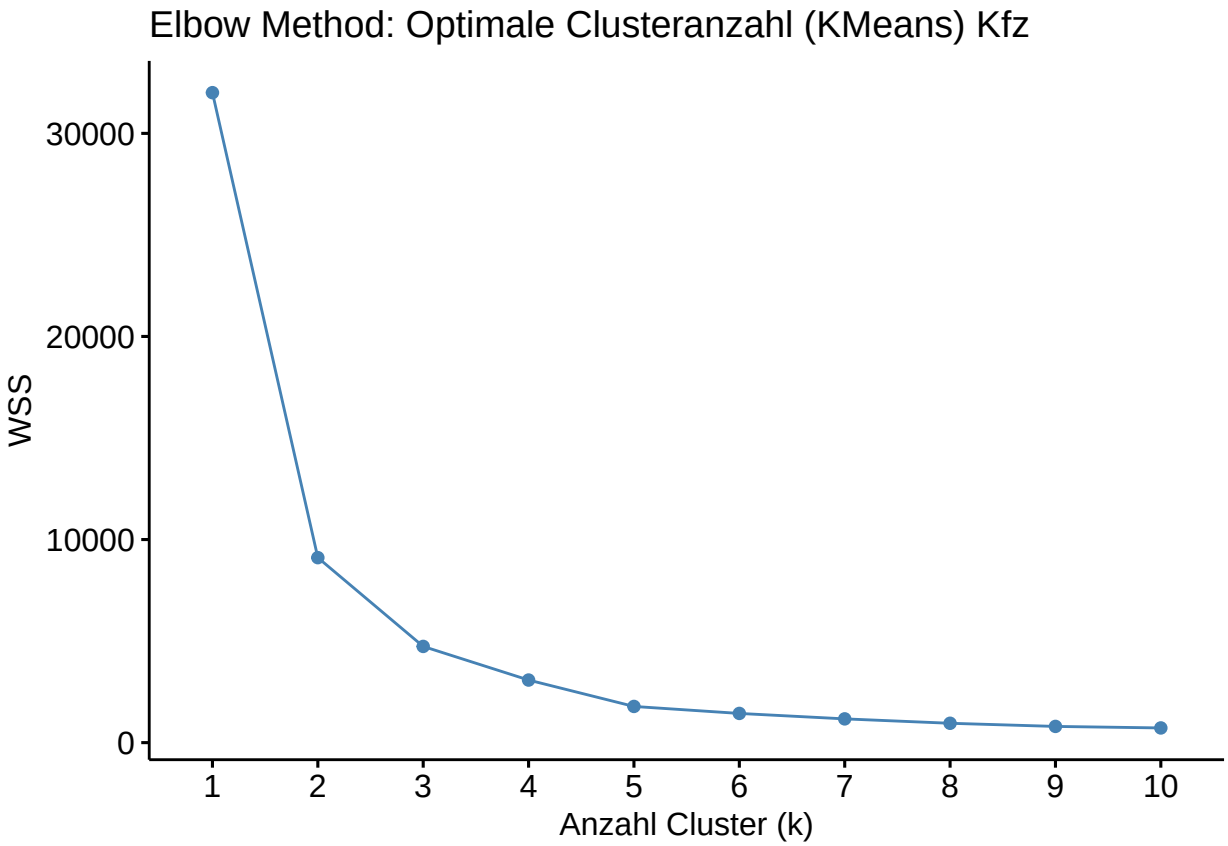
```
## Rows: 40418 Columns: 17
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr   (4): ZNAME, STRTYP, RINAME, FZTYP
## dbl  (11): ZNR, STRNR, DTVMS, DTVMF, DTVMO, DTVDD, DTVFR, DTVSA, DTVSF, TVMA...
## date  (2): DATUM, TVMAXT
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```r
# Nur Einträge mit RINAME == "Gesamt" => Auffahrten und Ausfahrten werden nicht berücksichtigt
gesamt <- verkehr %>% filter(RINAME == "Gesamt")

gesamt_kfz <- gesamt %>% filter(FZTYP == "Kfz")
gesamt_lkw <- gesamt %>% filter(FZTYP == "LkwÄ")

gesamt_kfz$ZNR <- as.integer(gesamt_kfz$ZNR)
gesamt_lkw$ZNR <- as.integer(gesamt_lkw$ZNR)
zaehlstellen$ZNR <- as.integer(zaehlstellen$ZNR)

# Mergen mit Koordinaten durch die logs
gesamt_kfz_geo <- inner_join(gesamt_kfz, zaehlstellen, by = "ZNR")
gesamt_lkw_geo <- inner_join(gesamt_lkw, zaehlstellen, by = "ZNR")
```

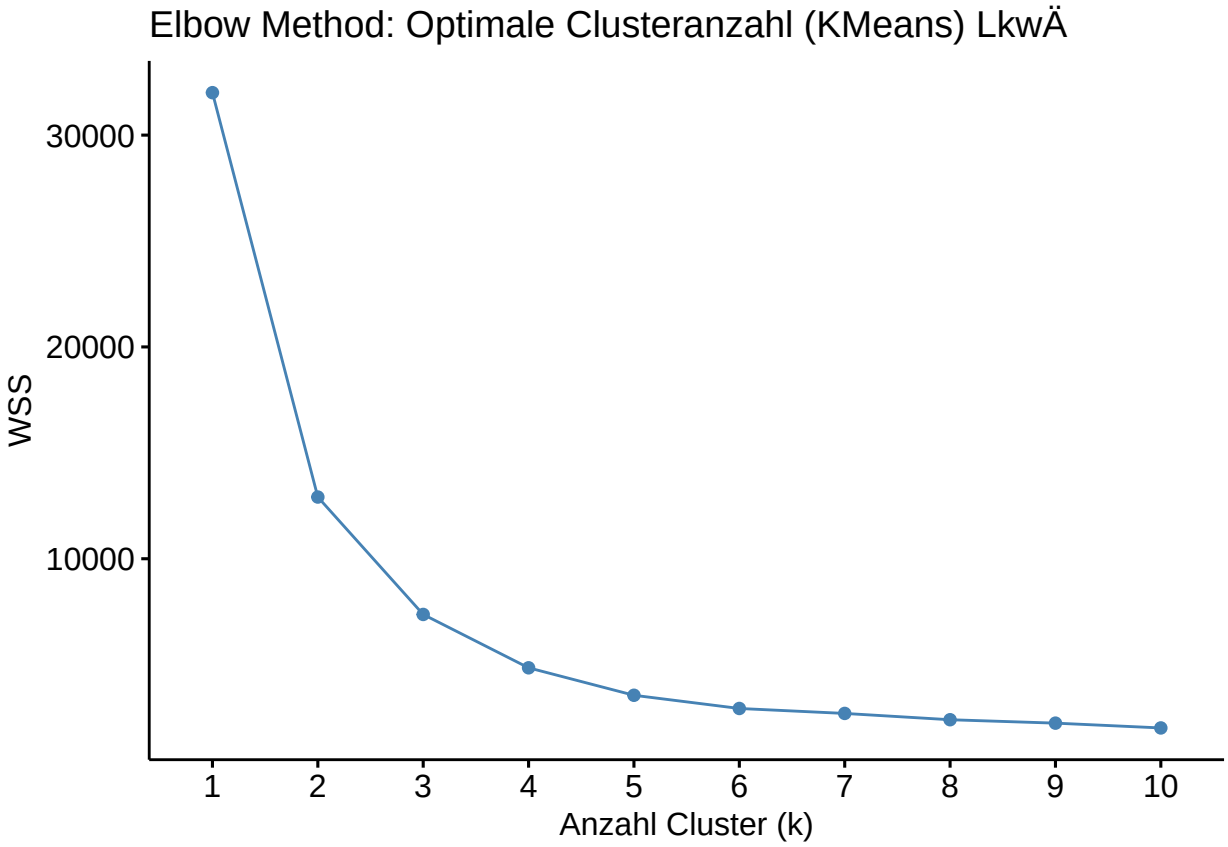
```r
ggplot() +
  geom_point(data = gesamt_kfz_geo, aes(x = as.numeric(LONGITUDE), y = as.numeric(LATITUDE)),
             color = "red", alpha = 0.6, size = 2) +
  labs(title = "Zählstellen Wien", x = "Längengrad", y = "Breitengrad") +
  theme_minimal()
```




```r
# Clusteranalyse (Features) gesamte Woche
cluster_features <- c("DTVMO", "DTVDD", "DTVFR", "DTVSA", "DTVSF")

# Daten bereinigen
kfz_data <- gesamt_kfz_geo %>%
  select(all_of(cluster_features)) %>%
  mutate_all(as.numeric) %>%
  drop_na()

lkw_data <- gesamt_lkw_geo %>%
  select(all_of(cluster_features)) %>%
  mutate_all(as.numeric) %>%
  drop_na()

kfz_scaled <- scale(kfz_data)
lkw_scaled <- scale(lkw_data)

# Elbow-Methode kfz
fviz_nbclust(kfz_scaled, kmeans, method = "wss") +
  labs(title = "Elbow Method: Optimale Clusteranzahl (KMeans) Kfz", x = "Anzahl Cluster (k)", y = "WSS")
```

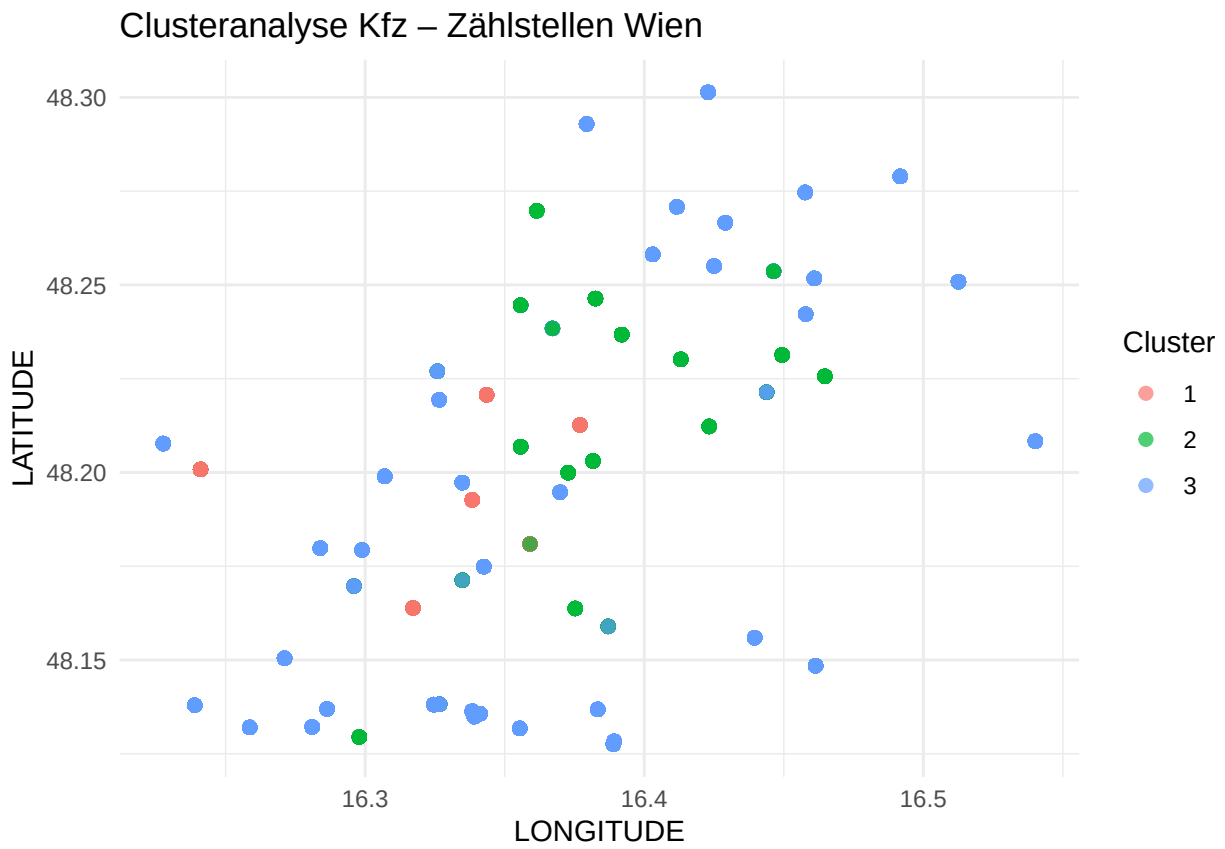



```
# Elbow-Methode lkw
fviz_nbclust(lkw_scaled, kmeans, method = "wss") +
  labs(title = "Elbow Method: Optimale Clusteranzahl (KMeans) LkwÄ", x = "Anzahl Cluster (k)", y = "WSS
```

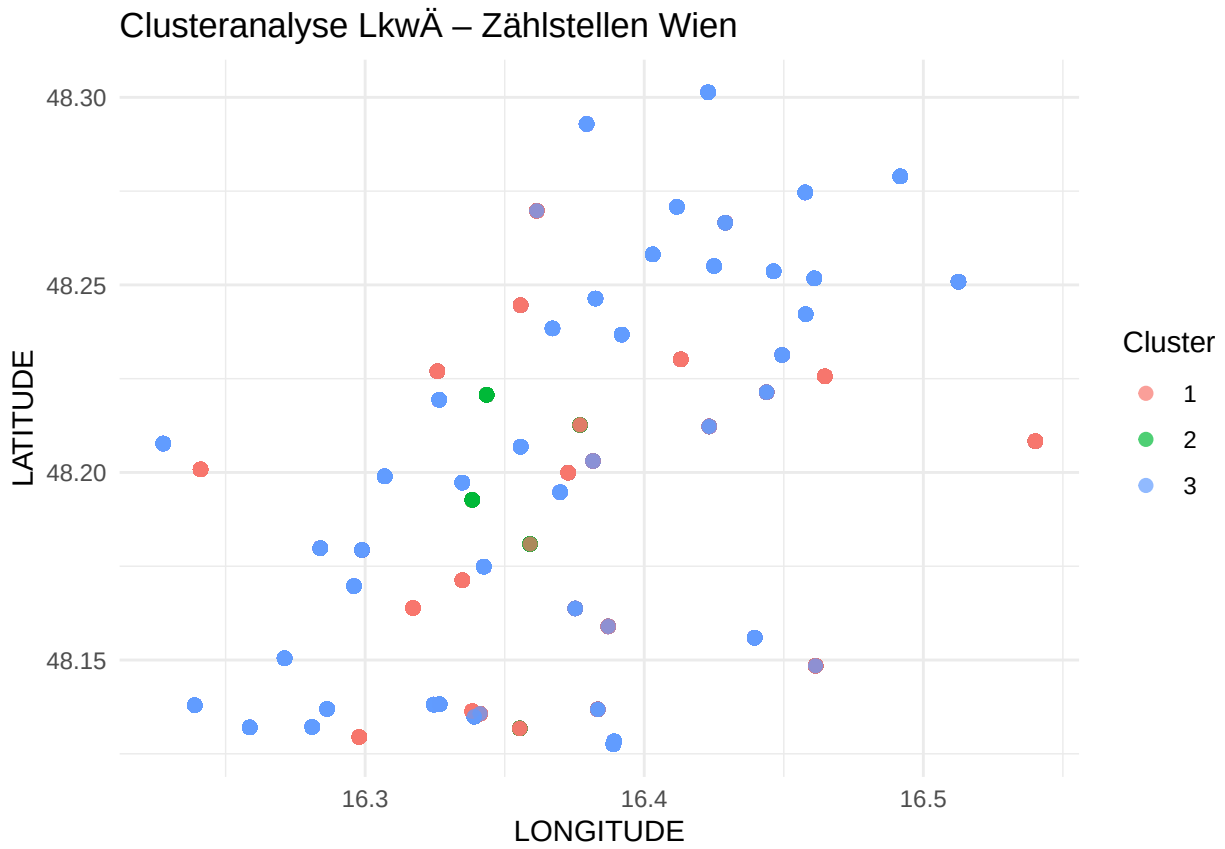



```r
# K-Means Cluster
kfz_cluster <- kmeans(kfz_scaled, centers = 3, nstart = 25)
lkw_cluster <- kmeans(lkw_scaled, centers = 3, nstart = 25)

# Cluster-Zuordnung hinzufügen
gesamt_kfz_geo_clean <- gesamt_kfz_geo[complete.cases(kfz_data), ]
gesamt_kfz_geo_clean$Cluster <- factor(kfz_cluster$cluster)

gesamt_lkw_geo_clean <- gesamt_lkw_geo[complete.cases(lkw_data), ]
gesamt_lkw_geo_clean$Cluster <- factor(lkw_cluster$cluster)
```

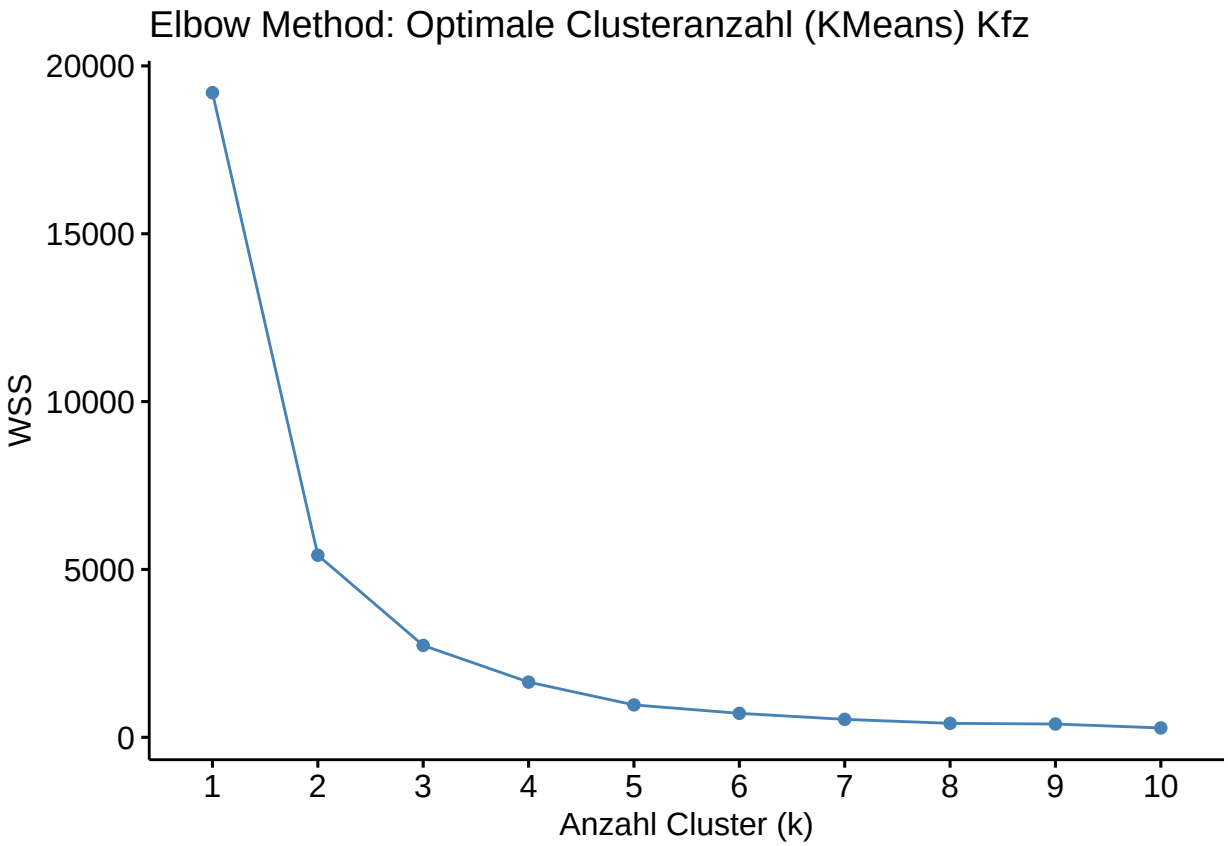
```r
# Plot Cluster Kfz
ggplot(gesamt_kfz_geo_clean, aes(x = LONGITUDE, y = LATITUDE, color = Cluster)) +
  geom_point(size = 2, alpha = 0.7) +
  labs(title = "Clusteranalyse Kfz - Zählstellen Wien") +
  theme_minimal()
```



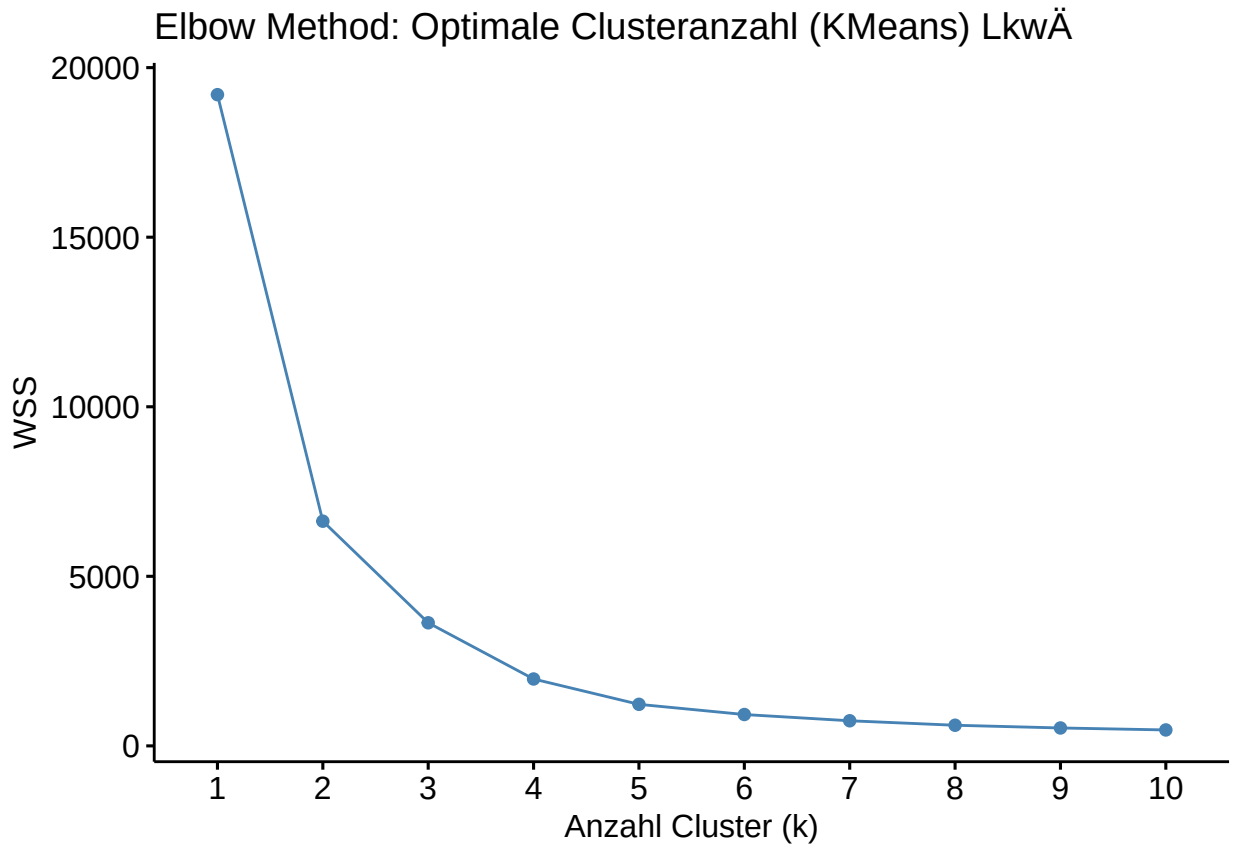

```r
# Plot Cluster Lkw
ggplot(gesamt_lkw_geo_clean, aes(x = LONGITUDE, y = LATITUDE, color = Cluster)) +
  geom_point(size = 2, alpha = 0.7) +
  labs(title = "Clusteranalyse LkwÄ - Zählstellen Wien") +
  theme_minimal()
```




```r
# Clusteranalyse (Features) der Tage Montag bis Freitag
cluster_features_mo_fr <- c("DTVMO", "DTVDD", "DTVFR")

# Daten bereinigen
kfz_data_mo_fr <- gesamt_kfz_geo %>%
  select(all_of(cluster_features_mo_fr)) %>%
  mutate_all(as.numeric) %>%
  drop_na()

lkw_data_mo_fr <- gesamt_lkw_geo %>%
  select(all_of(cluster_features_mo_fr)) %>%
  mutate_all(as.numeric) %>%
  drop_na()

kfz_scaled_mo_fr <- scale(kfz_data_mo_fr)
lkw_scaled_mo_fr <- scale(lkw_data_mo_fr)

# Elbow-Methode kfz
fviz_nbclust(kfz_scaled_mo_fr, kmeans, method = "wss") +
  labs(title = "Elbow Method: Optimale Clusteranzahl (KMeans) Kfz", x = "Anzahl Cluster (k)", y = "WSS"
```

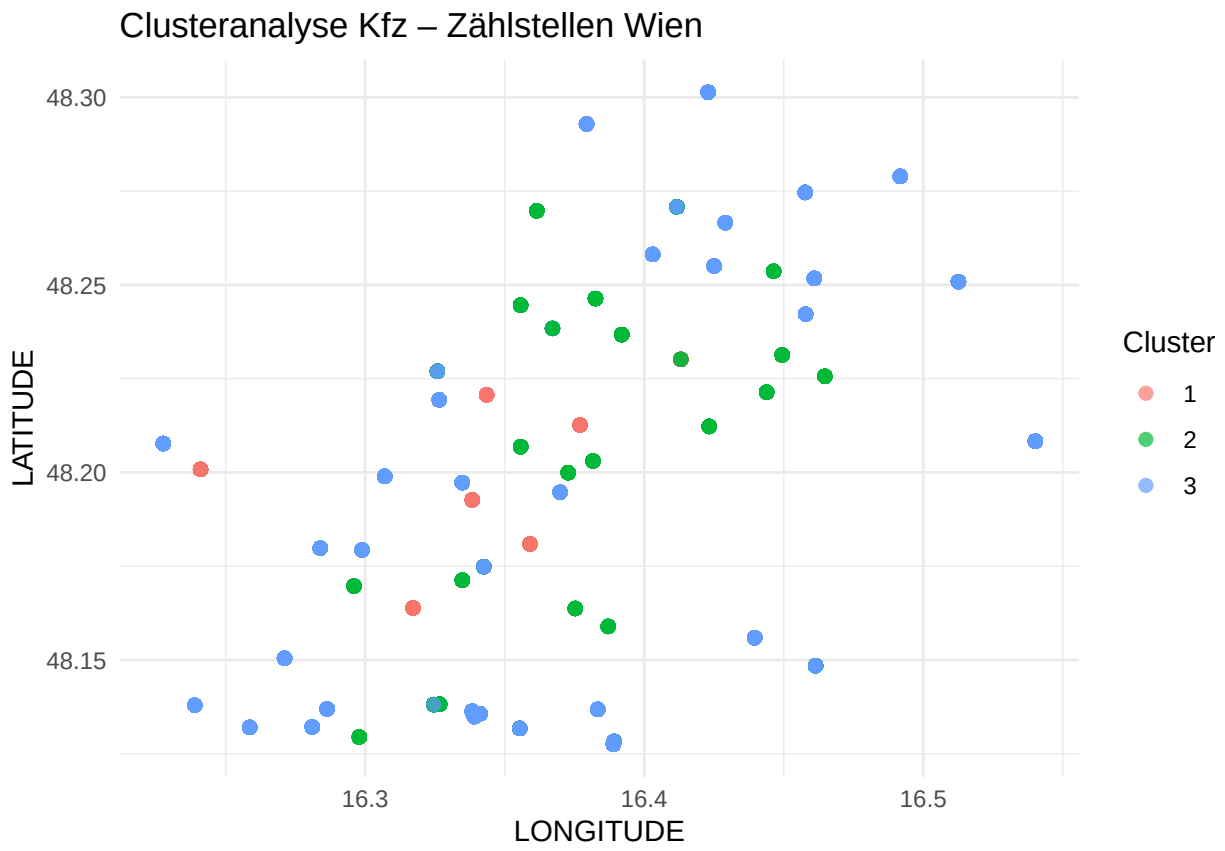



```
# Elbow-Methode lkw
fviz_nbclust(lkw_scaled_mo_fr, kmeans, method = "wss") +
  labs(title = "Elbow Method: Optimale Clusteranzahl (KMeans) LkwÄ", x = "Anzahl Cluster (k)", y = "WSS
```

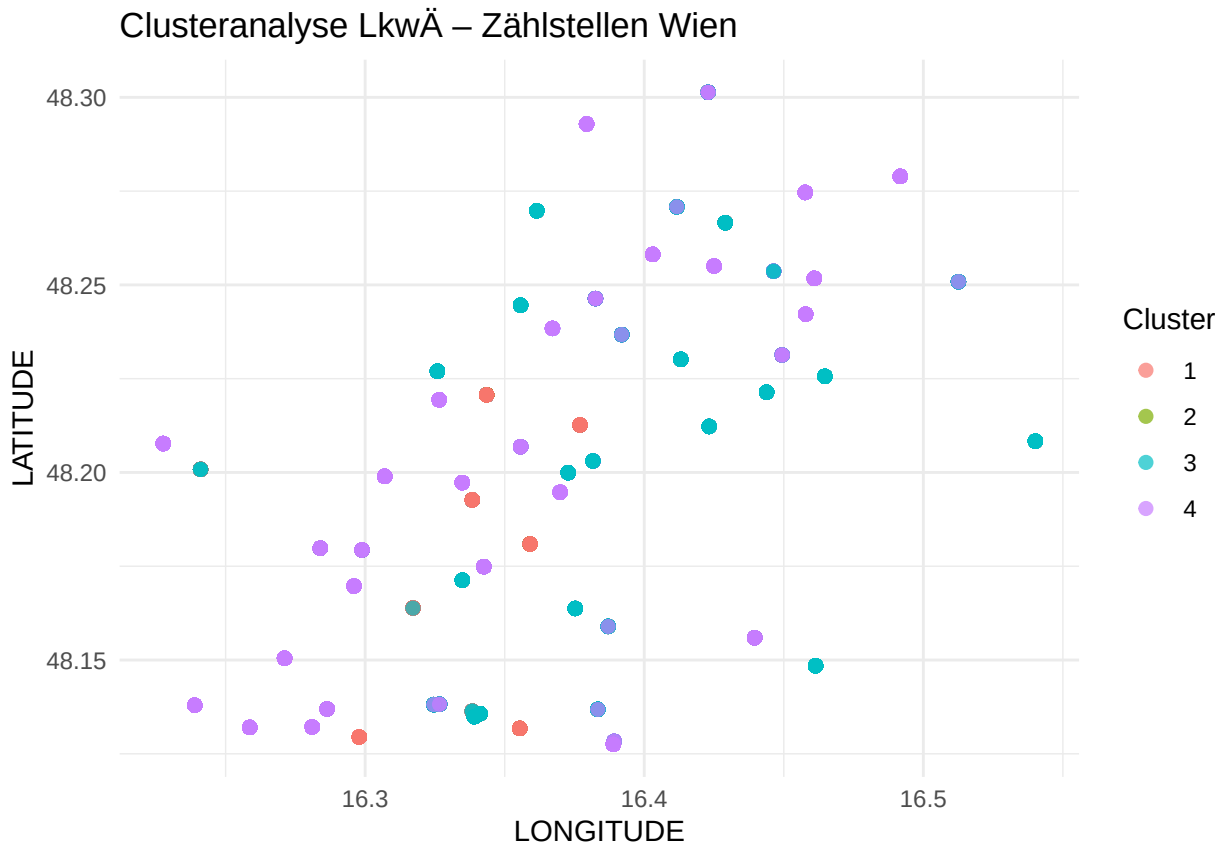



```r
# K-Means Cluster
kfz_cluster_mo_fr <- kmeans(kfz_scaled_mo_fr, centers = 3, nstart = 25)
lkw_cluster_mo_fr <- kmeans(lkw_scaled_mo_fr, centers = 4, nstart = 25)

# Cluster-Zuordnung hinzufügen
gesamt_kfz_geo_clean_mo_fr <- gesamt_kfz_geo[complete.cases(kfz_data_mo_fr), ]
gesamt_kfz_geo_clean_mo_fr$Cluster <- factor(kfz_cluster_mo_fr$cluster)

gesamt_lkw_geo_clean_mo_fr <- gesamt_lkw_geo[complete.cases(lkw_data_mo_fr), ]
gesamt_lkw_geo_clean_mo_fr$Cluster <- factor(lkw_cluster_mo_fr$cluster)
```

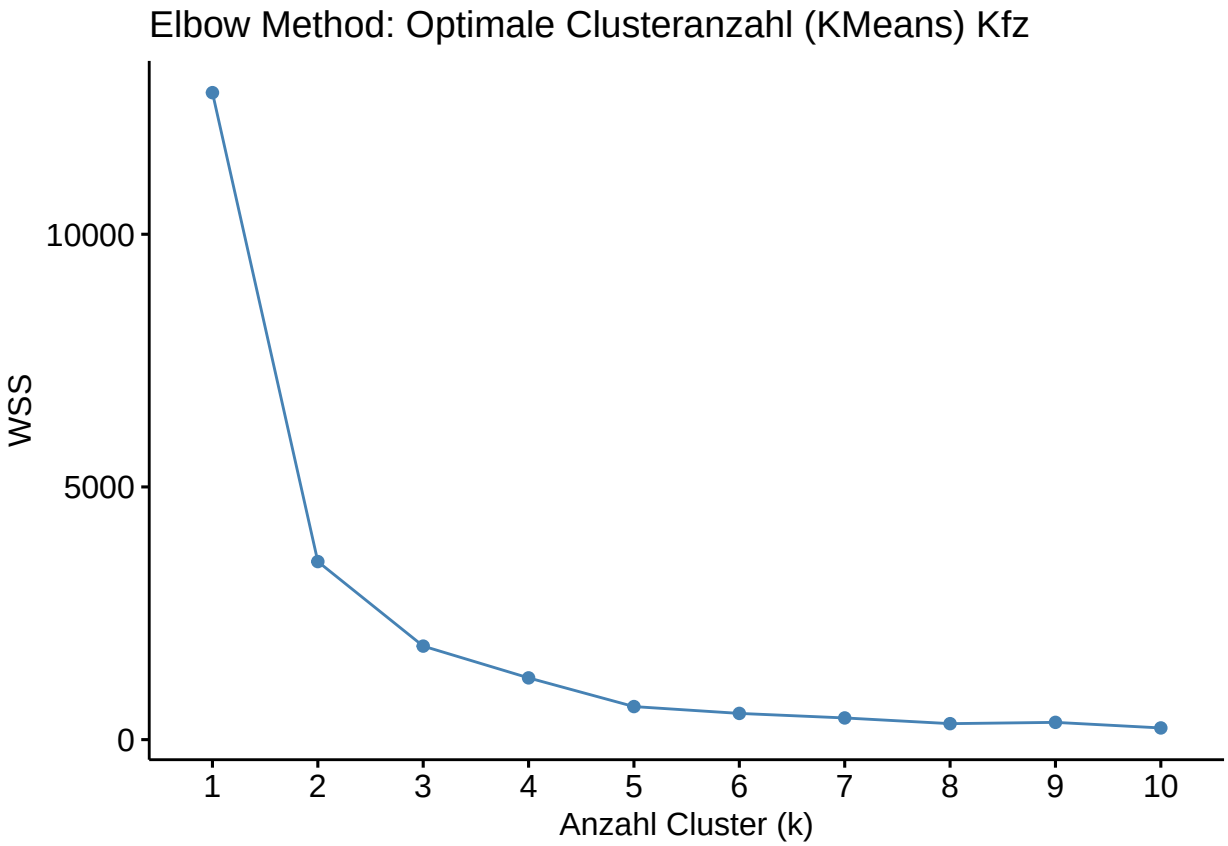
```r
# Plot Cluster Kfz
ggplot(gesamt_kfz_geo_clean_mo_fr, aes(x = LONGITUDE, y = LATITUDE, color = Cluster)) +
  geom_point(size = 2, alpha = 0.7) +
  labs(title = "Clusteranalyse Kfz - Zählstellen Wien") +
  theme_minimal()
```



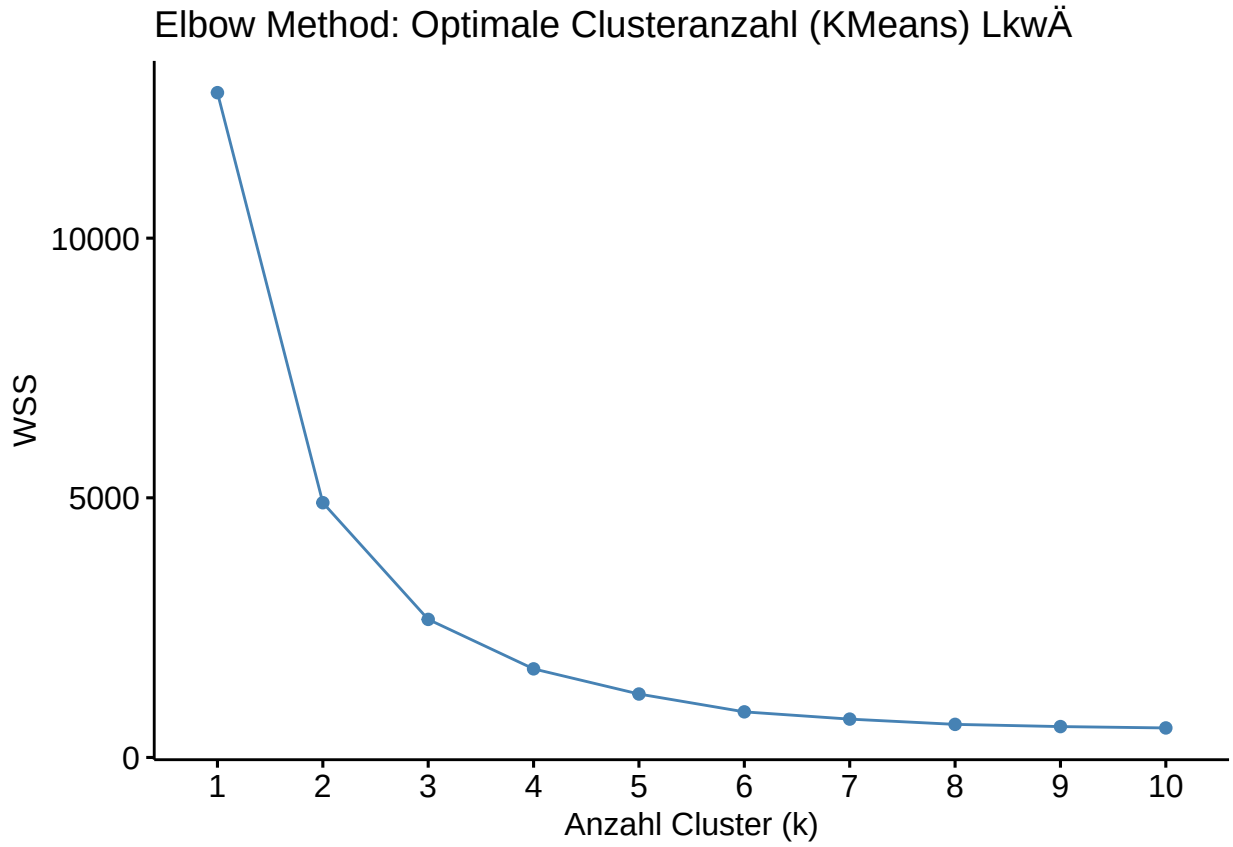

```r
# Plot Cluster Lkw
ggplot(gesamt_lkw_geo_clean_mo_fr, aes(x = LONGITUDE, y = LATITUDE, color = Cluster)) +
  geom_point(size = 2, alpha = 0.7) +
  labs(title = "Clusteranalyse LkwÄ - Zählstellen Wien") +
  theme_minimal()
```




```r
# Clusteranalyse (Features) nur Wochenende und Feiertage
cluster_features_we <- c("DTVSA", "DTVSF")

# Daten bereinigen
kfz_data_we <- gesamt_kfz_geo %>%
  select(all_of(cluster_features_we)) %>%
  mutate_all(as.numeric) %>%
  drop_na()

lkw_data_we <- gesamt_lkw_geo %>%
  select(all_of(cluster_features_we)) %>%
  mutate_all(as.numeric) %>%
  drop_na()

kfz_scaled_we <- scale(kfz_data_we)
lkw_scaled_we <- scale(lkw_data_we)

# Elbow-Methode kfz
fviz_nbclust(kfz_scaled_we, kmeans, method = "wss") +
  labs(title = "Elbow Method: Optimale Clusteranzahl (KMeans) Kfz", x = "Anzahl Cluster (k)", y = "WSS"
```

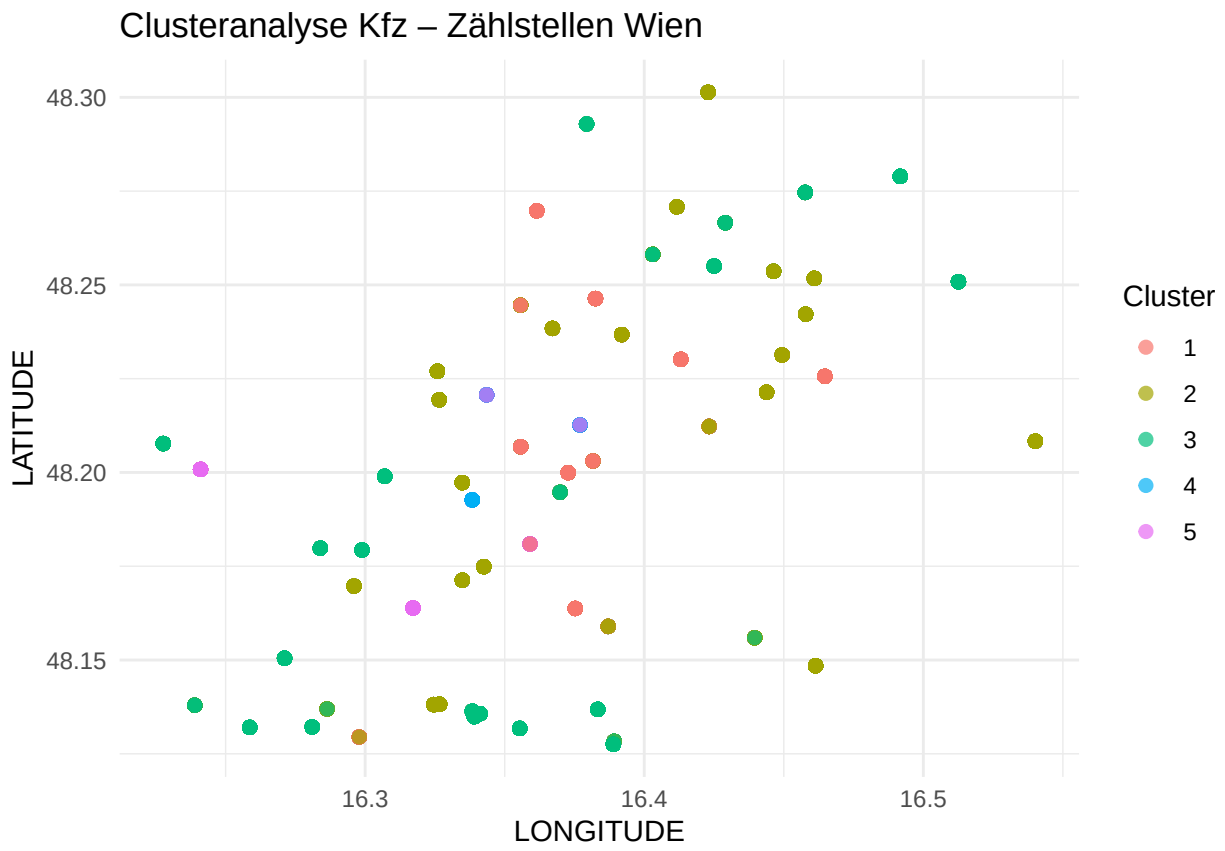



```
# Elbow-Methode lkw
fviz_nbclust(lkw_scaled_we, kmeans, method = "wss") +
  labs(title = "Elbow Method: Optimale Clusteranzahl (KMeans) LkwÄ", x = "Anzahl Cluster (k)", y = "WSS
```

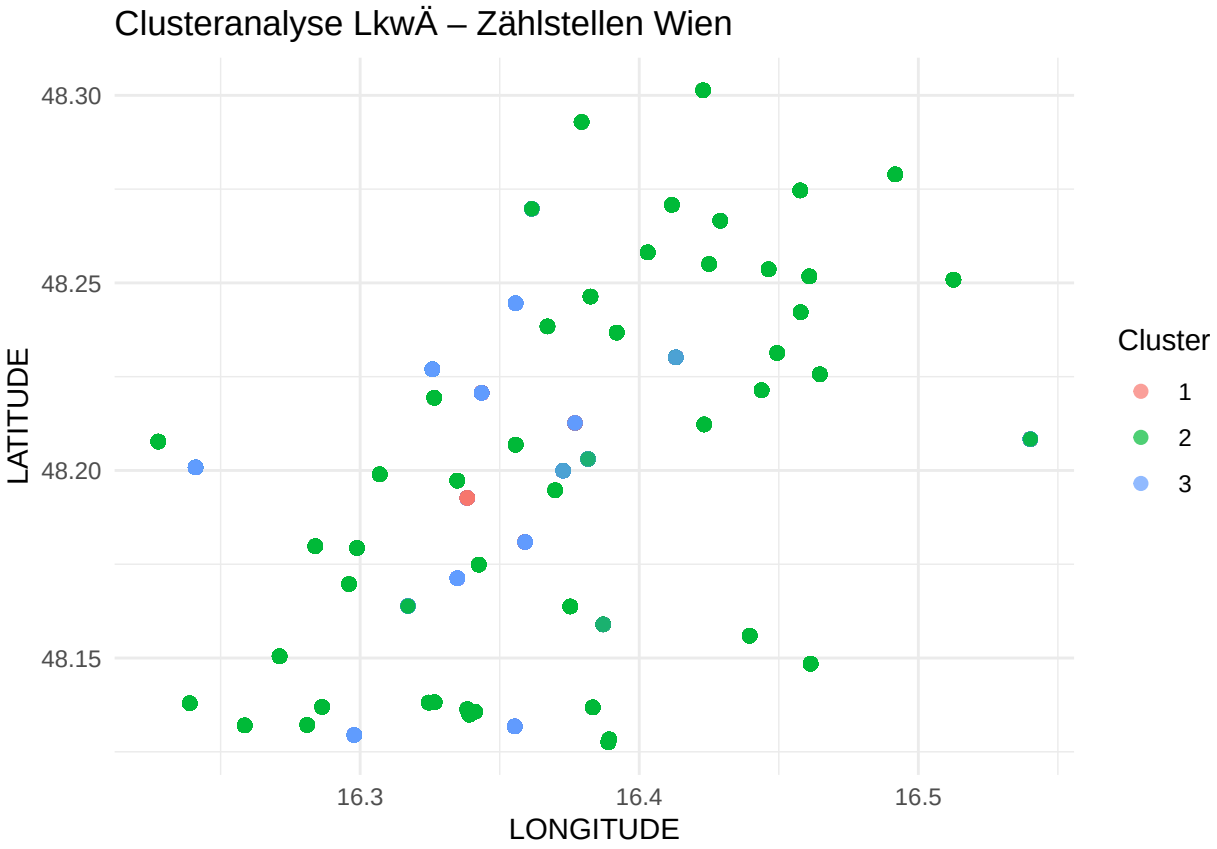



```r
# K-Means Cluster
kfz_cluster_we <- kmeans(kfz_scaled_we, centers = 5, nstart = 25)
lkw_cluster_we <- kmeans(lkw_scaled_we, centers = 3, nstart = 25)

# Cluster-Zuordnung hinzufügen
gesamt_kfz_geo_clean_we <- gesamt_kfz_geo[complete.cases(kfz_data_we), ]
gesamt_kfz_geo_clean_we$Cluster <- factor(kfz_cluster_we$cluster)

gesamt_lkw_geo_clean_we <- gesamt_lkw_geo[complete.cases(lkw_data_we), ]
gesamt_lkw_geo_clean_we$Cluster <- factor(lkw_cluster_we$cluster)
```

```r
# Plot Cluster Kfz
ggplot(gesamt_kfz_geo_clean_we, aes(x = LONGITUDE, y = LATITUDE, color = Cluster)) +
  geom_point(size = 2, alpha = 0.7) +
  labs(title = "Clusteranalyse Kfz - Zählstellen Wien") +
  theme_minimal()
```




```
# Plot Cluster Lkw
ggplot(gesamt_lkw_geo_clean_we, aes(x = LONGITUDE, y = LATITUDE, color = Cluster)) +
  geom_point(size = 2, alpha = 0.7) +
  labs(title = "Clusteranalyse LkwÄ - Zählstellen Wien") +
  theme_minimal()
```


Clusteranalyse LkwÄ – Zählstellen Wien
48.30
48.25
48.20
48.15
LATITUDE
16.3
16.4
16.5
LONGITUDE
Cluster
1
2
3