k-means

23.05.2025

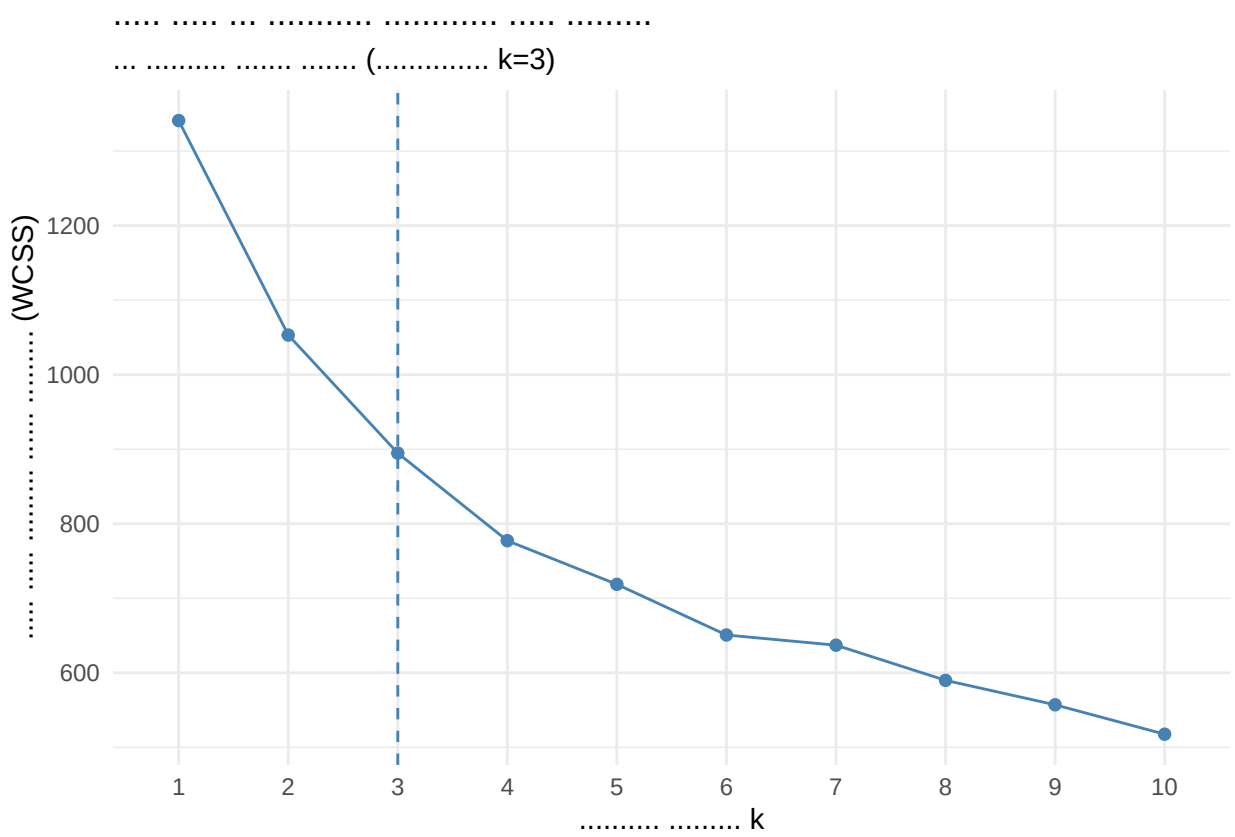
```r
gas_turbine_data <- read_csv("gas_turbine_power_dataset.csv", locale = locale(encoding = "UTF-8"))

#
turbine_numeric_features_for_clustering <- names(select(gas_turbine_data, where(is.numeric)))

head(gas_turbine_data) %>%
  knitr::kable(caption = "    6                            ")
```

Table 1: 6

AmbientTemperature	AmbientPressure	RelativeHumidity	FuelFlow_percent	CompressorOutletTemp	CompressorEfficiency	TurbineEfficiency	GasTurbineAge_years	MaintenanceCycle	PowerOutput_MW
16.2	102.54	23.6	25.54	17.0	91.89	92.04	5	Mid	158.06
33.5	99.20	57.2	21.17	32.3	86.32	87.00	18	Mid	116.48
27.0	98.72	57.8	18.48	27.1	88.22	90.07	10	Early	115.34
23.0	100.45	64.6	28.13	22.6	91.77	91.77	6	Early	174.07
9.7	102.93	70.8	12.22	12.0	90.64	93.64	1	Mid	86.16
9.7	99.21	88.3	19.85	10.2	88.39	94.78	5	Mid	125.51

```r
#             (                   )
scaled_data <- gas_turbine_data %>%
  select(where(is.numeric)) %>%
  scale()
```

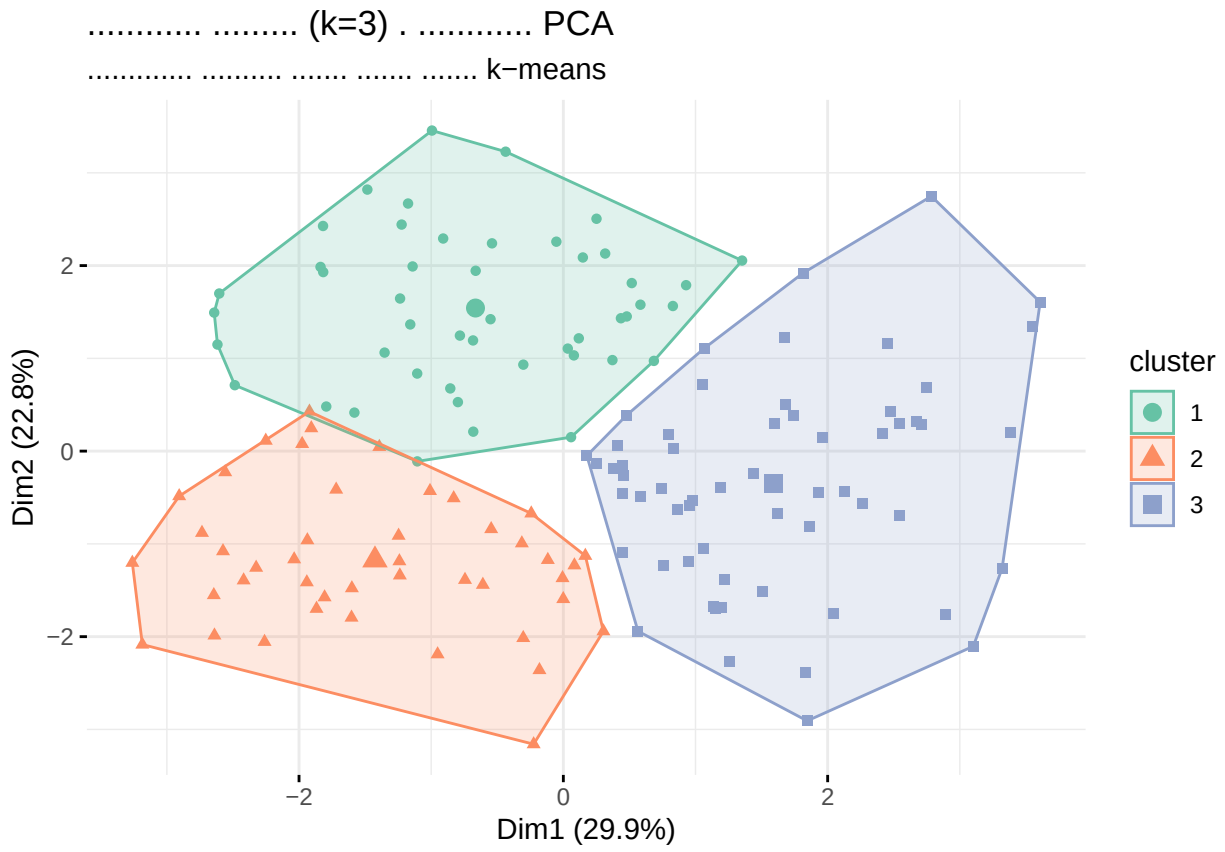
```r
fviz_nbclust(scaled_data, kmeans, method = "wss") +
  geom_vline(xintercept = 3, linetype = 2, color = "steelblue") +
  labs(title = "                                ",
       subtitle = "                  (         k=3)",
       x = "            k",
       y = "                        (WCSS)") +
  theme_minimal()
```



k-means

```
chosen_k <- 3
set.seed(123)
km_res <- kmeans(scaled_data, centers = chosen_k, nstart = 25)

gas_turbine_clustered <- gas_turbine_data %>%
  mutate(Cluster = as.factor(km_res$cluster))
```

```
fviz_cluster(km_res, data = scaled_data,
             palette = "Set2",
             geom = "point",
             ellipse.type = "convex",
             ggtheme = theme_minimal(),
             main = paste("             (k=", chosen_k, ")          PCA", sep="")
             ) +
  labs(subtitle = "                            k-means")
```



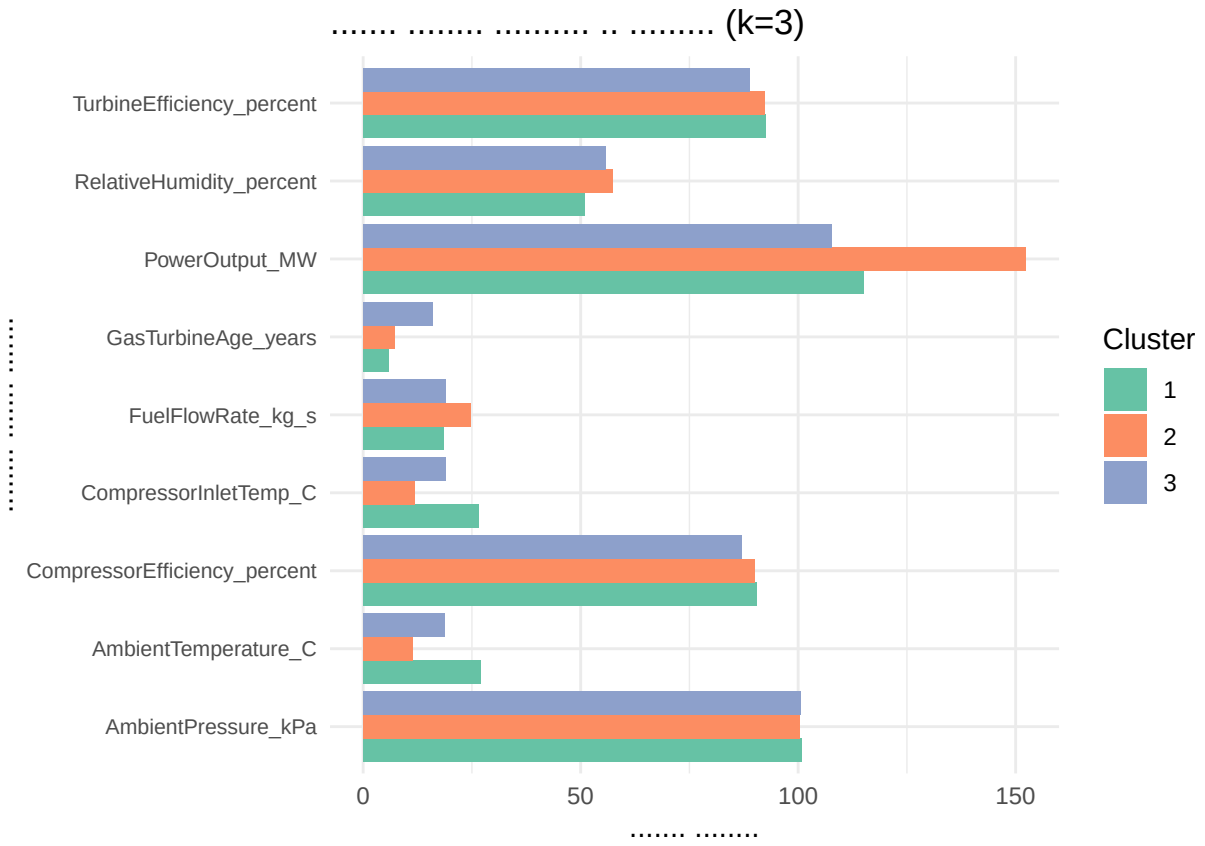

```
cluster_summary_means <- gas_turbine_clustered %>%
  group_by(Cluster) %>%
  summarise(across(where(is.numeric), mean, na.rm = TRUE))

cluster_summary_means %>%
  knitr::kable(caption = paste("                                    (k=", chosen_k, ")", sep=""), digits =
```

Table 2: (k=3)

Cluster	Ambient_Temperature_C	Ambient_Pressure_kPa	Relative_Humidity_pct	Fuel_Flow_kg_s	Compressor_Pressure_Ratio	Compressor_Efficiency_pct	Turbine_Efficiency_pct	Gas_Turbine_Temperature_C	Power_Output_MW
1	26.91	100.80	51.03	18.61	26.53	90.41	92.62	5.94	115.12
2	11.41	100.32	57.38	24.68	11.75	90.08	92.36	7.20	152.36
3	18.84	100.62	55.69	18.91	18.97	87.09	88.94	15.88	107.62

```
features_to_plot_means <- c(turbine_numeric_features_for_clustering, "PowerOutput_MW")
```

```r
cluster_summary_means %>%
  select(Cluster, all_of(features_to_plot_means)) %>%
  pivot_longer(cols = -Cluster, names_to = "     ", values_to = "     _     ") %>%
  ggplot(aes(x =      , y =      _      , fill = Cluster)) +
    geom_col(position = "dodge") +
    scale_fill_brewer(palette = "Set2") +
    coord_flip() +
    labs(title = paste("                              (k=", chosen_k, ")", sep=""),
         x = "                ",
         y = "          ") +
    theme_minimal() +
    theme(axis.text.y = element_text(size=8))
```



,

. ,

. .