

Traffic FInal Code

Erin Xu

2025-11-14

Setup

```
library(openxlsx)
library(tidyverse)

## Warning: package 'ggplot2' was built under R version 4.5.2

## Warning: package 'readr' was built under R version 4.5.2

## -- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
## v dplyr      1.1.4      v readr      2.1.6
## v forcats    1.0.1      v stringr    1.5.2
## v ggplot2    4.0.1      v tibble     3.3.0
## v lubridate  1.9.4      v tidyr      1.3.1
## v purrr      1.1.0
## -- 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

library(fastICA)
library(ggplot2)
library(ggvenn)
library(tidyr)
library(cluster)
file <- "traffic.xlsx"
sheet_names <- getSheetNames(file)
num_sheets <- length(sheet_names)
print(sheet_names)
```

```
## [1] "Loc1" "Loc2" "Loc3" "Loc4" "Loc5" "Loc6" "Loc7" "Loc8" "Loc9"
## [10] "Loc10" "Loc11" "Loc12" "Loc13" "Loc14" "Loc15" "Loc16" "Loc17" "Loc18"
## [19] "Loc19" "Loc20" "Loc21" "Loc22" "Loc23" "Loc24" "Loc25" "Loc26"
```

```
print(num_sheets)
```

```
## [1] 26
```

```
set.seed(67)
df <- lapply(sheet_names, function(sheet) {
  as.matrix(read.xlsx(file, sheet = sheet, colNames = TRUE))
})

names(df) <- sheet_names

# Check one location
str(df[[1]])
```

```
## num [1:288, 1:384] 138 129 121 115 109 104 100 97 95 93 ...
## - attr(*, "dimnames")=List of 2
## ..$ : chr [1:288] "1" "2" "3" "4" ...
## ..$ : chr [1:384] "WkDay-1" "WkDay-2" "WkDay-3" "WkDay-4" ...
```

Helpers for anomalies

```
choose_k_pca <- function(pca, threshold = 0.90) {
  var_expl <- pca$sdev^2
  var_expl <- var_expl / sum(var_expl)
  cumvar <- cumsum(var_expl)
  k <- which(cumvar >= threshold)[1]
  return(k)
}

find_anomalies_from_scores <- function(score_mat, multiplier = 1.5) {
  # score_mat: rows = days, cols = components
  n_days <- nrow(score_mat)
  is_outlier <- rep(FALSE, n_days)

  for (j in seq_len(ncol(score_mat))) {
    x <- score_mat[, j]
    stats <- boxplot.stats(x, coef = multiplier)
```

```

    out_idx <- which(x %in% stats$out)
    is_outlier[out_idx] <- TRUE
  }

  which(is_outlier) # returns row indices of anomalous days
}

analyze_location <- function(loc_name,
                             X,
                             var_expl_threshold = 0.90,
                             max_factors       = 3,
                             anomaly_coef      = 1.5,
                             do_ica            = FALSE,
                             n_ica_comp        = 3) {

  message("\nProcessing location: ", loc_name)

  # ---- 1. Transpose matrix so rows = days, columns = timepoints ----
  X_t <- t(X)
  X_centered <- scale(X_t, center = TRUE, scale = FALSE)

  # ---- 2. PCA ----
  pca <- prcomp(X_centered, center = FALSE, scale. = FALSE)
  k_pca <- choose_k_pca(pca, threshold = var_expl_threshold)

  pca_scores <- pca$x[, 1:k_pca, drop = FALSE]
  pca_loadings <- pca$rotation[, 1:k_pca, drop = FALSE]

  # ---- 3. PCA-based anomaly detection ----
  pca_anom_idx <- find_anomalies_from_scores(pca_scores, multiplier = anomaly_coef)
  pca_anom_days <- rownames(pca_scores)[pca_anom_idx]

  # ---- 4. Factor Analysis (on PCA scores for numerical stability) ----
  fa_model <- NULL
  fa_scores <- NULL
  fa_loadings <- NULL
  fa_anom_days <- character(0)

  # Use fewer factors for interpretability (typically 2-5)
  n_factors <- min(max_factors, k_pca, 5)

  if (n_factors >= 1) {
    # Use FA on PCA scores to avoid singularity issues
    n_pcs_for_fa <- min(50, k_pca)
  }
}

```

```

n_factors_fa <- min(n_factors, n_pcs_for_fa - 1)

if (n_factors_fa >= 1 && n_pcs_for_fa >= 3) {
  fa_model <- tryCatch(
    factanal(pca_scores[, 1:n_pcs_for_fa, drop = FALSE],
             factors = n_factors_fa,
             scores = "regression",
             rotation = "varimax"),
    error = function(e) {
      message(" FA failed for ", loc_name, ": ", e$message)
      return(NULL)
    }
  )

  if (!is.null(fa_model)) {
    fa_scores <- fa_model$scores
    # Map loadings back to original time-of-day space
    fa_loadings <- pca_loadings[, 1:n_pcs_for_fa] %*% fa_model$loadings[, , drop = F
    fa_anom_idx <- find_anomalies_from_scores(fa_scores, multiplier = anomaly_coef)
    fa_anom_days <- rownames(fa_scores)[fa_anom_idx]
    message(" FA succeeded on PCA scores")
  }
} else {
  message(" FA skipped for ", loc_name, ": insufficient PCs")
}
} else {
  message(" FA skipped for ", loc_name, ": insufficient factors")
}

# ---ica on PCA scores---

ica_result <- NULL
ica_anom_days <- character(0)
ica_scores <- NULL
ica_loadings <- NULL

if (do_ica && k_pca >= 2) {
  n_ica_to_use <- min(n_ica_comp, k_pca)

  ica_result <- tryCatch(
    fastICA(pca_scores, n.comp = n_ica_to_use, method = "C"),
    error = function(e) {
      message(" ICA failed for ", loc_name, ": ", e$message)
      return(NULL)
    }
  )
}

```

```

    }
  )

  if (!is.null(ica_result)) {
    # ICA scores (strength of each source per day)
    ica_scores <- ica_result$S

    # Map ICA loadings back to original time-of-day space
    # ica_result$A maps PCA space -> ICA sources
    # pca_loadings maps original space -> PCA space
    ica_loadings <- pca_loadings[, 1:n_ica_to_use, drop = FALSE] %*% ica_result$A

    if (!is.null(ica_scores) && nrow(ica_scores) > 0) {
      message(" ICA succeeded with ", ncol(ica_scores), " sources.")

      # Ensure day names - use pca_scores rownames which are correct!
      if (is.null(rownames(ica_scores))) {
        rownames(ica_scores) <- rownames(pca_scores)
      }

      # Calculate ICA anomalies
      ica_anom_idx <- find_anomalies_from_scores(ica_scores, multiplier = anomaly_coef)
      ica_anom_days <- rownames(ica_scores)[ica_anom_idx]

    } else {
      message(" ICA failed to produce valid scores for ", loc_name)
      ica_result <- NULL
      ica_scores <- NULL
      ica_loadings <- NULL
    }
  }
}

# ---- 6. Return results (Now includes all ICA components) ----
list(
  location      = loc_name,
  pca           = pca,
  k_pca        = k_pca,
  pca_scores    = pca_scores,
  pca_loadings  = pca_loadings,
  pca_anom_days = pca_anom_days,
  fa_model      = fa_model,
  fa_scores     = fa_scores,
  fa_loadings   = fa_loadings,

```

```

    fa_anom_days = fa_anom_days,
    ica           = ica_result,
    # ADDED: Explicitly return ICA components
    ica_scores    = ica_scores,
    ica_loadings  = ica_loadings,
    ica_anom_days = ica_anom_days
  )
}

# ---- Run for all locations ----
location_results <- lapply(names(df), function(loc_name) {
  analyze_location(
    loc_name = loc_name,
    X         = df[[loc_name]],
    var_expl_threshold = 0.90,
    max_factors        = 3,
    anomaly_coef       = 1.5,
    do_ica             = TRUE
  )
})

##
## Processing location: Loc1

##   FA failed for Loc1: 3 factors are too many for 5 variables

##   ICA succeeded with 3 sources.

##
## Processing location: Loc2

##   FA succeeded on PCA scores

##   ICA succeeded with 3 sources.

##
## Processing location: Loc3

##   FA failed for Loc3: 3 factors are too many for 5 variables

##   ICA succeeded with 3 sources.

```

```
##
## Processing location: Loc4

##   FA succeeded on PCA scores

##   ICA succeeded with 3 sources.

##
## Processing location: Loc5

##   FA succeeded on PCA scores

##   ICA succeeded with 3 sources.

##
## Processing location: Loc6

##   FA succeeded on PCA scores

##   ICA succeeded with 3 sources.

##
## Processing location: Loc7

##   FA skipped for Loc7: insufficient PCs

##   ICA succeeded with 2 sources.

##
## Processing location: Loc8

##   FA succeeded on PCA scores

##   ICA succeeded with 3 sources.

##
## Processing location: Loc9

##   FA failed for Loc9: 3 factors are too many for 4 variables

##   ICA succeeded with 3 sources.
```

```
##
## Processing location: Loc10

## FA failed for Loc10: 3 factors are too many for 5 variables

## ICA succeeded with 3 sources.

##
## Processing location: Loc11

## FA succeeded on PCA scores

## ICA succeeded with 3 sources.

##
## Processing location: Loc12

## FA succeeded on PCA scores

## ICA succeeded with 3 sources.

##
## Processing location: Loc13

## FA succeeded on PCA scores

## ICA succeeded with 3 sources.

##
## Processing location: Loc14

## FA succeeded on PCA scores

## ICA succeeded with 3 sources.

##
## Processing location: Loc15

## FA succeeded on PCA scores

## ICA succeeded with 3 sources.
```



```
##
## Processing location: Loc16

## FA succeeded on PCA scores

## ICA succeeded with 3 sources.

##
## Processing location: Loc17

## FA succeeded on PCA scores

## ICA succeeded with 3 sources.

##
## Processing location: Loc18

## FA failed for Loc18: 3 factors are too many for 5 variables

## ICA succeeded with 3 sources.

##
## Processing location: Loc19

## FA succeeded on PCA scores

## ICA succeeded with 3 sources.

##
## Processing location: Loc20

## FA succeeded on PCA scores

## ICA succeeded with 3 sources.

##
## Processing location: Loc21

## FA failed for Loc21: 3 factors are too many for 5 variables

## ICA succeeded with 3 sources.
```

```

##
## Processing location: Loc22

## FA succeeded on PCA scores

## ICA succeeded with 3 sources.

##
## Processing location: Loc23

## FA succeeded on PCA scores

## ICA succeeded with 3 sources.

##
## Processing location: Loc24

## FA failed for Loc24: 2 factors are too many for 3 variables

## ICA succeeded with 3 sources.

##
## Processing location: Loc25

## FA succeeded on PCA scores

## ICA succeeded with 3 sources.

##
## Processing location: Loc26

## FA skipped for Loc26: insufficient PCs

## ICA succeeded with 2 sources.

names(location_results) <- names(df)

```

List of anomalies


```

## =====
## Location: Loc4
## = = = = =
## PCA: retained 7 components
## PCA anomalies ( 20 ): WkDay-5, WkDay-27, WkDay-29, WkDay-34, WkDay-73, WkDay-91, WkDay-117, WkDay-138, WkDay-159, WkDay-180, WkDay-201, WkDay-222, WkDay-241, WkDay-261, WkDay-282, WkDay-303, WkDay-324, WkDay-345, WkDay-366, WkDay-387, WkDay-408, WkDay-429, WkDay-450, WkDay-471, WkDay-492, WkDay-513, WkDay-534, WkDay-555, WkDay-576, WkDay-597, WkDay-618, WkDay-639, WkDay-660, WkDay-681, WkDay-702, WkDay-723, WkDay-744, WkDay-765, WkDay-786, WkDay-807, WkDay-828, WkDay-849, WkDay-870, WkDay-891, WkDay-912, WkDay-933, WkDay-954, WkDay-975, WkDay-996, WkDay-1017, WkDay-1038, WkDay-1059, WkDay-1080, WkDay-1101, WkDay-1122, WkDay-1143, WkDay-1164, WkDay-1185, WkDay-1206, WkDay-1227, WkDay-1248, WkDay-1269, WkDay-1290, WkDay-1311, WkDay-1332, WkDay-1353, WkDay-1374, WkDay-1395, WkDay-1416, WkDay-1437, WkDay-1458, WkDay-1479, WkDay-1500, WkDay-1521, WkDay-1542, WkDay-1563, WkDay-1584, WkDay-1605, WkDay-1626, WkDay-1647, WkDay-1668, WkDay-1689, WkDay-1710, WkDay-1731, WkDay-1752, WkDay-1773, WkDay-1794, WkDay-1815, WkDay-1836, WkDay-1857, WkDay-1878, WkDay-1899, WkDay-1920, WkDay-1941, WkDay-1962, WkDay-1983, WkDay-2004, WkDay-2025, WkDay-2046, WkDay-2067, WkDay-2088, WkDay-2109, WkDay-2130, WkDay-2151, WkDay-2172, WkDay-2193, WkDay-2214, WkDay-2235, WkDay-2256, WkDay-2277, WkDay-2298, WkDay-2319, WkDay-2340, WkDay-2361, WkDay-2382, WkDay-2403, WkDay-2424, WkDay-2445, WkDay-2466, WkDay-2487, WkDay-2508, WkDay-2529, WkDay-2550, WkDay-2571, WkDay-2592, WkDay-2613, WkDay-2634, WkDay-2655, WkDay-2676, WkDay-2697, WkDay-2718, WkDay-2739, WkDay-2760, WkDay-2781, WkDay-2802, WkDay-2823, WkDay-2844, WkDay-2865, WkDay-2886, WkDay-2907, WkDay-2928, WkDay-2949, WkDay-2970, WkDay-2991, WkDay-3012, WkDay-3033, WkDay-3054, WkDay-3075, WkDay-3096, WkDay-3117, WkDay-3138, WkDay-3159, WkDay-3180, WkDay-3201, WkDay-3222, WkDay-3243, WkDay-3264, WkDay-3285, WkDay-3306, WkDay-3327, WkDay-3348, WkDay-3369, WkDay-3390, WkDay-3411, WkDay-3432, WkDay-3453, WkDay-3474, WkDay-3495, WkDay-3516, WkDay-3537, WkDay-3558, WkDay-3579, WkDay-3600, WkDay-3621, WkDay-3642, WkDay-3663, WkDay-3684, WkDay-3705, WkDay-3726, WkDay-3747, WkDay-3768, WkDay-3789, WkDay-3810, WkDay-3831, WkDay-3852, WkDay-3873, WkDay-3894, WkDay-3915, WkDay-3936, WkDay-3957, WkDay-3978, WkDay-3999, WkDay-4020, WkDay-4041, WkDay-4062, WkDay-4083, WkDay-4104, WkDay-4125, WkDay-4146, WkDay-4167, WkDay-4188, WkDay-4209, WkDay-4230, WkDay-4251, WkDay-4272, WkDay-4293, WkDay-4314, WkDay-4335, WkDay-4356, WkDay-4377, WkDay-4398, WkDay-4419, WkDay-4440, WkDay-4461, WkDay-4482, WkDay-4503, WkDay-4524, WkDay-4545, WkDay-4566, WkDay-4587, WkDay-4608, WkDay-4629, WkDay-4650, WkDay-4671, WkDay-4692, WkDay-4713, WkDay-4734, WkDay-4755, WkDay-4776, WkDay-4797, WkDay-4818, WkDay-4839, WkDay-4860, WkDay-4881, WkDay-4902, WkDay-4923, WkDay-4944, WkDay-4965, WkDay-4986, WkDay-5007, WkDay-5028, WkDay-5049, WkDay-5070, WkDay-5091, WkDay-5112, WkDay-5133, WkDay-5154, WkDay-5175, WkDay-5196, WkDay-5217, WkDay-5238, WkDay-5259, WkDay-5280, WkDay-5301, WkDay-5322, WkDay-5343, WkDay-5364, WkDay-5385, WkDay-5406, WkDay-5427, WkDay-5448, WkDay-5469, WkDay-5490, WkDay-5511, WkDay-5532, WkDay-5553, WkDay-5574, WkDay-5595, WkDay-5616, WkDay-5637, WkDay-5658, WkDay-5679, WkDay-5700, WkDay-5721, WkDay-5742, WkDay-5763, WkDay-5784, WkDay-5805, WkDay-5826, WkDay-5847, WkDay-5868, WkDay-5889, WkDay-5910, WkDay-5931, WkDay-5952, WkDay-5973, WkDay-5994, WkDay-6015, WkDay-6036, WkDay-6057, WkDay-6078, WkDay-6099, WkDay-6120, WkDay-6141, WkDay-6162, WkDay-6183, WkDay-6204, WkDay-6225, WkDay-6246, WkDay-6267, WkDay-6288, WkDay-6309, WkDay-6330, WkDay-6351, WkDay-6372, WkDay-6393, WkDay-6414, WkDay-6435, WkDay-6456, WkDay-6477, WkDay-6498, WkDay-6519, WkDay-6540, WkDay-6561, WkDay-6582, WkDay-6603, WkDay-6624, WkDay-6645, WkDay-6666, WkDay-6687, WkDay-6708, WkDay-6729, WkDay-6750, WkDay-6771, WkDay-6792, WkDay-6813, WkDay-6834, WkDay-6855, WkDay-6876, WkDay-6897, WkDay-6918, WkDay-6939, WkDay-6960, WkDay-6981, WkDay-7002, WkDay-7023, WkDay-7044, WkDay-7065, WkDay-7086, WkDay-7107, WkDay-7128, WkDay-7149, WkDay-7170, WkDay-7191, WkDay-7212, WkDay-7233, WkDay-7254, WkDay-7275, WkDay-7296, WkDay-7317, WkDay-7338, WkDay-7359, WkDay-7380, WkDay-7401, WkDay-7422, WkDay-7443, WkDay-7464, WkDay-7485, WkDay-7506, WkDay-7527, WkDay-7548, WkDay-7569, WkDay-7590, WkDay-7611, WkDay-7632, WkDay-7653, WkDay-7674, WkDay-7695, WkDay-7716, WkDay-7737, WkDay-7758, WkDay-7779, WkDay-7800, WkDay-7821, WkDay-7842, WkDay-7863, WkDay-7884, WkDay-7905, WkDay-7926, WkDay-7947, WkDay-7968, WkDay-7989, WkDay-8010, WkDay-8031, WkDay-8052, WkDay-8073, WkDay-8094, WkDay-8115, WkDay-8136, WkDay-8157, WkDay-8178, WkDay-8199, WkDay-8220, WkDay-8241, WkDay-8262, WkDay-8283, WkDay-8304, WkDay-8325, WkDay-8346, WkDay-8367, WkDay-8388, WkDay-8409, WkDay-8430, WkDay-8451, WkDay-8472, WkDay-8493, WkDay-8514, WkDay-8535, WkDay-8556, WkDay-8577, WkDay-8598, WkDay-8619, WkDay-8640, WkDay-8661, WkDay-8682, WkDay-8703, WkDay-8724, WkDay-8745, WkDay-8766, WkDay-8787, WkDay-8808, WkDay-8829, WkDay-8850, WkDay-8871, WkDay-8892, WkDay-8913, WkDay-8934, WkDay-8955, WkDay-8976, WkDay-8997, WkDay-9018, WkDay-9039, WkDay-9060, WkDay-9081, WkDay-9102, WkDay-9123, WkDay-9144, WkDay-9165, WkDay-9186, WkDay-9207, WkDay-9228, WkDay-9249, WkDay-9270, WkDay-9291, WkDay-9312, WkDay-9333, WkDay-9354, WkDay-9375, WkDay-9396, WkDay-9417, WkDay-9438, WkDay-9459, WkDay-9480, WkDay-9501, WkDay-9522, WkDay-9543, WkDay-9564, WkDay-9585, WkDay-9606, WkDay-9627, WkDay-9648, WkDay-9669, WkDay-9690, WkDay-9711, WkDay-9732, WkDay-9753, WkDay-9774, WkDay-9795, WkDay-9816, WkDay-9837, WkDay-9858, WkDay-9879, WkDay-9900, WkDay-9921, WkDay-9942, WkDay-9963, WkDay-9984, WkDay-10005, WkDay-10026, WkDay-10047, WkDay-10068, WkDay-10089, WkDay-10110, WkDay-10131, WkDay-10152, WkDay-10173, WkDay-10194, WkDay-10215, WkDay-10236, WkDay-10257, WkDay-10278, WkDay-10299, WkDay-10320, WkDay-10341, WkDay-10362, WkDay-10383, WkDay-10404, WkDay-10425, WkDay-10446, WkDay-10467, WkDay-10488, WkDay-10509, WkDay-10530, WkDay-10551, WkDay-10572, WkDay-10593, WkDay-10614, WkDay-10635, WkDay-10656, WkDay-10677, WkDay-10698, WkDay-10719, WkDay-10740, WkDay-10761, WkDay-10782, WkDay-10803, WkDay-10824, WkDay-10845, WkDay-10866, WkDay-10887, WkDay-10908, WkDay-10929, WkDay-10950, WkDay-10971, WkDay-10992, WkDay-11013, WkDay-11034, WkDay-11055, WkDay-11076, WkDay-11097, WkDay-11118, WkDay-11139, WkDay-11160, WkDay-11181, WkDay-11202, WkDay-11223, WkDay-11244, WkDay-11265, WkDay-11286, WkDay-11307, WkDay-11328, WkDay-11349, WkDay-11370, WkDay-11391, WkDay-11412, WkDay-11433, WkDay-11454, WkDay-11475, WkDay-11496, WkDay-11517, WkDay-11538, WkDay-11559, WkDay-11580, WkDay-11601, WkDay-11622, WkDay-11643, WkDay-11664, WkDay-11685, WkDay-11706, WkDay-11727, WkDay-11748, WkDay-11769, WkDay-11790, WkDay-11811, WkDay-11832, WkDay-11853, WkDay-11874, WkDay-11895, WkDay-11916, WkDay-11937, WkDay-11958, WkDay-11979, WkDay-11999, WkDay-12020, WkDay-12041, WkDay-12062, WkDay-12083, WkDay-12104, WkDay-12125, WkDay-12146, WkDay-12167, WkDay-12188, WkDay-12209, WkDay-12230, WkDay-12251, WkDay-12272, WkDay-12293, WkDay-12314, WkDay-12335, WkDay-12356, WkDay-12377, WkDay-12398, WkDay-12419, WkDay-12440, WkDay-12461, WkDay-12482, WkDay-12503, WkDay-12524, WkDay-12545, WkDay-12566, WkDay-12587, WkDay-12608, WkDay-12629, WkDay-12650, WkDay-12671, WkDay-12692, WkDay-12713, WkDay-12734, WkDay-12755, WkDay-12776, WkDay-12797, WkDay-12818, WkDay-12839, WkDay-12860, WkDay-12881, WkDay-12902, WkDay-12923, WkDay-12944, WkDay-12965, WkDay-12986, WkDay-13007, WkDay-13028, WkDay-13049, WkDay-13070, WkDay-13091, WkDay-13112, WkDay-13133, WkDay-13154, WkDay-13175, WkDay-13196, WkDay-13217, WkDay-13238, WkDay-13259, WkDay-13280, WkDay-13301, WkDay-13322, WkDay-13343, WkDay-13364, WkDay-13385, WkDay-13406, WkDay-13427, WkDay-13448, WkDay-13469, WkDay-13490, WkDay-13511, WkDay-13532, WkDay-13553, WkDay-13574, WkDay-13595, WkDay-13616, WkDay-13637, WkDay-13658, WkDay-13679, WkDay-13700, WkDay-13721, WkDay-13742, WkDay-13763, WkDay-13784, WkDay-13805, WkDay-13826, WkDay-13847, WkDay-13868, WkDay-13889, WkDay-13910, WkDay-13931, WkDay-13952, WkDay-13973, WkDay-13994, WkDay-14015, WkDay-14036, WkDay-14057, WkDay-14078, WkDay-14099, WkDay-14120, WkDay-14141, WkDay-14162, WkDay-14183, WkDay-14204, WkDay-14225, WkDay-14246, WkDay-14267, WkDay-14288, WkDay-14309, WkDay-14330, WkDay-14351, WkDay-14372, WkDay-14393, WkDay-14414, WkDay-14435, WkDay-14456, WkDay-14477, WkDay-14498, WkDay-14519, WkDay-14540, WkDay-14561, WkDay-14582, WkDay-14603, WkDay-14624, WkDay-14645, WkDay-14666, WkDay-14687, WkDay-14708, WkDay-14729, WkDay-14750, WkDay-14771, WkDay-14792, WkDay-14813, WkDay-14834, WkDay-14855, WkDay-14876, WkDay-14897, WkDay-14918, WkDay-14939, WkDay-14960, WkDay-14981, WkDay-15002, WkDay-15023, WkDay-15044, WkDay-15065, WkDay-15086, WkDay-15107, WkDay-15128, WkDay-15149, WkDay-15170, WkDay-15191, WkDay-15212, WkDay-15233, WkDay-15254, WkDay-15275, WkDay-15296, WkDay-15317, WkDay-15338, WkDay-15359, WkDay-15380, WkDay-15401, WkDay-15422, WkDay-15443, WkDay-15464, WkDay-15485, WkDay-15506, WkDay-15527, WkDay-15548, WkDay-15569, WkDay-15590, WkDay-15611, WkDay-15632, WkDay-15653, WkDay-15674, WkDay-15695, WkDay-15716, WkDay-15737, WkDay-15758, WkDay-15779, WkDay-15800, WkDay-15821, WkDay-15842, WkDay-15863, WkDay-15884, WkDay-15905, WkDay-15926, WkDay-15947, WkDay-15968, WkDay-15989, WkDay-16010, WkDay-16031, WkDay-16052, WkDay-16073, WkDay-16094, WkDay-16115, WkDay-16136, WkDay-16157, WkDay-16178, WkDay-16199, WkDay-16220, WkDay-16241, WkDay-16262, WkDay-16283, WkDay-16304, WkDay-16325, WkDay-16346, WkDay-16367, WkDay-16388, WkDay-16409, WkDay-16430, WkDay-16451, WkDay-16472, WkDay-16493, WkDay-16514, WkDay-16535, WkDay-16556, WkDay-16577, WkDay-16598, WkDay-16619, WkDay-16640, WkDay-16661, WkDay-16682, WkDay-16703, WkDay-16724, WkDay-16745, WkDay-16766, WkDay-16787, WkDay-16808, WkDay-16829, WkDay-16850, WkDay-16871, WkDay-16892, WkDay-16913, WkDay-16934, WkDay-16955, WkDay-16976, WkDay-16997, WkDay-17018, WkDay-17039, WkDay-17060, WkDay-17081, WkDay-17102, WkDay-17123, WkDay-17144, WkDay-17165, WkDay-17186, WkDay-17207, WkDay-17228, WkDay-17249, WkDay-17270, WkDay-17291, WkDay-17312, WkDay-17333, WkDay-17354, WkDay-17375, WkDay-17396, WkDay-17417, WkDay-17438, WkDay-17459, WkDay-17480, WkDay-17501, WkDay-17522, WkDay-17543, WkDay-17564, WkDay-17585, WkDay-17606, WkDay-17627, WkDay-17648, WkDay-17669, WkDay-17690, WkDay-17711, WkDay-17732, WkDay-17753, WkDay-17774, WkDay-17795, WkDay-17816, WkDay-17837, WkDay-17858, WkDay-17879, WkDay-17900, WkDay-17921, WkDay-17942, WkDay-17963, WkDay-17984, WkDay-18005, WkDay-18026, WkDay-18047, WkDay-18068, WkDay-18089, WkDay-18110, WkDay-18131, WkDay-18152, WkDay-18173, WkDay-18194, WkDay-18215, WkDay-18236, WkDay-18257, WkDay-18278, WkDay-18299, WkDay-18320, WkDay-18341, WkDay-18362, WkDay-18383, WkDay-18404, WkDay-18425, WkDay-18446, WkDay-18467, WkDay-18488, WkDay-18509, WkDay-18530, WkDay-18551, WkDay-18572, WkDay-18593, WkDay-18614, WkDay-18635, WkDay-18656, WkDay-18677, WkDay-18698, WkDay-18719, WkDay-18740, WkDay-18761, WkDay-18782, WkDay-18803, WkDay-18824, WkDay-18845, WkDay-18866, WkDay-18887, WkDay-18908, WkDay-18929, WkDay-18950, WkDay-18971, WkDay-18992, WkDay-19013, WkDay-19034, WkDay-19055, WkDay-19076, WkDay-19097, WkDay-19118, WkDay-19139, WkDay-19160, WkDay-19181, WkDay-19202, WkDay-19223, WkDay-19244, WkDay-19265, WkDay-19286, WkDay-19307, WkDay-19328, WkDay-19349, WkDay-19370, WkDay-19391, WkDay-19412, WkDay-19433, WkDay-19454, WkDay-19475, WkDay-19496, WkDay-19517, WkDay-19538, WkDay-19559, WkDay-19580, WkDay-19601, WkDay-19622, WkDay-19643, WkDay-19664, WkDay-19685, WkDay-19706, WkDay-19727, WkDay-19748, WkDay-19769, WkDay-19790, WkDay-19811, WkDay-19832, WkDay-19853, WkDay-19874, WkDay-19895, WkDay-19916, WkDay-19937, WkDay-19958, WkDay-19979, WkDay-19999, WkDay-20020, WkDay-20041, WkDay-20062, WkDay-20083, WkDay-20104, WkDay-20125, WkDay-20146, WkDay-20167, WkDay-20188, WkDay-20209, WkDay-20230, WkDay-20251, WkDay-20272, WkDay-20293, WkDay-20314, WkDay-20335, WkDay-20356, WkDay-20377, WkDay-20398, WkDay-20419, WkDay-20440, WkDay-20461, WkDay-20482, WkDay-20503, WkDay-20524, WkDay-20545, WkDay-20566, WkDay-20587, WkDay-20608, WkDay-20629, WkDay-20650, WkDay-20671, WkDay-20692, WkDay-20713, WkDay-20734, WkDay-20755, WkDay-20776, WkDay-20797, WkDay-20818, WkDay-20839, WkDay-20860, WkDay-20881, WkDay-20902, WkDay-20923, WkDay-20944, WkDay-20965, WkDay-20986, WkDay-21007, WkDay-21028, WkDay-21049, WkDay-21070, WkDay-21091, WkDay-21112, WkDay-21133, WkDay-21154, WkDay-21175, WkDay-21196, WkDay-21217, WkDay-21238, WkDay-21259, WkDay-21280, WkDay-21301, WkDay-21322, WkDay-21343, WkDay-21364, WkDay-21385, WkDay-21406, WkDay-21427, WkDay-21448, WkDay-21469, WkDay-21490, WkDay-21511, WkDay-21532, WkDay-21553, WkDay-21574, WkDay-21595, WkDay-21616, WkDay-21637, WkDay-21658, WkDay-21679, WkDay-21700, WkDay-21721, WkDay-21742, WkDay-21763, WkDay-21784, WkDay-21805, WkDay-21826, WkDay-21847, WkDay-21868, WkDay-21889, WkDay-21910, WkDay-21931, WkDay-21952, WkDay-21973, WkDay-21994, WkDay-22015, WkDay-22036, WkDay-22057, WkDay-22078, WkDay-22099, WkDay-22120, WkDay-22141, WkDay-22162, WkDay-22183, WkDay-22204, WkDay-22225, WkDay-22246, WkDay-22267, WkDay-22288, WkDay-22309, WkDay-22330, WkDay-22351, WkDay-22372, WkDay-22393, WkDay-22414, WkDay-22435, WkDay-22456, WkDay-22477, WkDay-22498, WkDay-22519, WkDay-22540, WkDay-22561, WkDay-22582, WkDay-22603, WkDay-22624, WkDay-22645, WkDay-22666, WkDay-22687, WkDay-22708, WkDay-22729, WkDay-22750, WkDay-22771, WkDay-22792, WkDay-22813, WkDay-22834, WkDay-22855, WkDay-22876, WkDay-22897, WkDay-22918, WkDay-22939, WkDay-22960, WkDay-22981, WkDay-23002, WkDay-23023, WkDay-23044, WkDay-23065, WkDay-23086, WkDay-23107, WkDay-23128, WkDay-23149, WkDay-23170, WkDay-23191, WkDay-23212, WkDay-23233, WkDay-23254, WkDay-23275, WkDay-23296, WkDay-23317, WkDay-23338, WkDay-23359, WkDay-23380, WkDay-23401, WkDay-23422, WkDay-23443, WkDay-23464, WkDay-23485, WkDay-23506, WkDay-23527, WkDay-23548, WkDay-23569, WkDay-23590, WkDay-23611, WkDay-23632, WkDay-23653, WkDay-23674, WkDay-23695, WkDay-23716, WkDay-23737, WkDay-23758, WkDay-23779, WkDay-23800, WkDay-23821, WkDay-23842, WkDay-23863, WkDay-23884, WkDay-23905, WkDay-23926, WkDay-23947, WkDay-23968, WkDay-23989, WkDay-24010, WkDay-24031, WkDay-24052, WkDay-24073, WkDay-24094, WkDay-24115, WkDay-24136, WkDay-24157, WkDay-24178, WkDay-24199, WkDay-24220, WkDay-24241, WkDay-24262, WkDay-24283, WkDay-24304, WkDay-24325, WkDay-24346, WkDay-24367, WkDay-24388, WkDay-24409, WkDay-24430, WkDay-24451, WkDay-24472, WkDay-24493, WkDay-24514, WkDay-24535, WkDay-24556, WkDay-24577, WkDay-24598, WkDay-24619, WkDay-24640, WkDay-24661, WkDay-24682, WkDay-24703, WkDay-24724, WkDay-24745, WkDay-24766, WkDay-24787, WkDay-24808, WkDay-24829, WkDay-24850, WkDay-24871, WkDay-24892, WkDay-24913, WkDay-24934, WkDay-24955, WkDay-24976, WkDay-24997, WkDay-25018, WkDay-25039, WkDay-25060, WkDay-25081, WkDay-25102, WkDay-25123, WkDay-25144, WkDay-25165, WkDay-25186, WkDay-25207, WkDay-25228, WkDay-25249, WkDay-25270, WkDay-25291, WkDay-25312, WkDay-25333, WkDay-25354, WkDay-25375, WkDay-25396, WkDay-25417, WkDay-25438, WkDay-25459, WkDay-25480, WkDay-25501, WkDay-25522, WkDay-25543, WkDay-25564, WkDay-25585, WkDay-25606, WkDay-25627, WkDay-25648, WkDay-25669, WkDay-25690, WkDay-25711, WkDay-25732, WkDay-25753, WkDay-25774, WkDay-25795, WkDay-25816, WkDay-25837, WkDay-25858, WkDay-25879, WkDay-25900, WkDay-25921, WkDay-25942, WkDay-25963, WkDay-25984, WkDay-26005, WkDay-26026, WkDay-26047, WkDay-26068, WkDay-26089, WkDay-26110, WkDay-26131, WkDay-26152, WkDay-26173, WkDay-26194, WkDay-26215, WkDay-26236, WkDay-26257, WkDay-26278, WkDay-26299, WkDay-26320, WkDay-26341, WkDay-26362, WkDay-26383, WkDay-26404, WkDay-26425, WkDay-26446, WkDay-26467, WkDay-26488, WkDay-26509, WkDay-26530, WkDay-26551, WkDay-26572, WkDay-26593, WkDay-26614, WkDay-26635, WkDay-26656, WkDay-26677, WkDay-26698, WkDay-26719, WkDay-26740, WkDay-26761, WkDay-26782, WkDay-26803, WkDay-26824, WkDay-26845, WkDay-26866, WkDay-26887, WkDay-26908, WkDay-26929, WkDay-26950, WkDay-2
```

```

## =====
## Location: Loc9
## = = = = =
## PCA: retained 4 components
## PCA anomalies ( 14 ): WkDay-24, WkDay-40, WkDay-50, WkDay-60, WkDay-127, WkDay-248, W
## FA anomalies ( 0 ):
## ICA anomalies ( 12 ): WkDay-24, WkDay-37, WkDay-50, WkDay-206, WkDay-248, WkDay-289,
##
##
## =====
## Location: Loc10
## = = = = =
## PCA: retained 5 components
## PCA anomalies ( 13 ): WkDay-46, WkDay-62, WkDay-86, WkDay-100, WkDay-169, WkDay-195,
## FA anomalies ( 0 ):
## ICA anomalies ( 7 ): WkDay-9, WkDay-10, WkDay-51, WkDay-62, WkDay-169, WkDay-171, WkD
##
##
## =====
## Location: Loc11
## = = = = =
## PCA: retained 6 components
## PCA anomalies ( 24 ): WkDay-1, WkDay-4, WkDay-7, WkDay-13, WkDay-15, WkDay-26, WkDay-
## FA anomalies ( 10 ): WkDay-1, WkDay-4, WkDay-26, WkDay-54, WkDay-66, WkDay-76, WkDay-
## ICA anomalies ( 9 ): WkDay-1, WkDay-2, WkDay-13, WkDay-14, WkDay-54, WkDay-222, WkDay
##
##
## =====
## Location: Loc12
## = = = = =
## PCA: retained 8 components
## PCA anomalies ( 23 ): WkDay-1, WkDay-4, WkDay-7, WkDay-19, WkDay-55, WkDay-131, WkDay
## FA anomalies ( 14 ): WkDay-7, WkDay-19, WkDay-55, WkDay-70, WkDay-77, WkDay-152, WkDa
## ICA anomalies ( 19 ): WkDay-1, WkDay-7, WkDay-19, WkDay-70, WkDay-77, WkDay-121, WkDa
##
##
## =====
## Location: Loc13
## = = = = =
## PCA: retained 9 components
## PCA anomalies ( 20 ): WkDay-1, WkDay-25, WkDay-31, WkDay-33, WkDay-34, WkDay-98, WkDa
## FA anomalies ( 9 ): WkDay-1, WkDay-50, WkDay-98, WkDay-139, WkDay-269, WkDay-271, WkD
## ICA anomalies ( 17 ): WkDay-1, WkDay-13, WkDay-16, WkDay-27, WkDay-37, WkDay-98, WkDa
##
##

```

```

## =====
## Location: Loc14
## = = = = =
## PCA: retained 8 components
## PCA anomalies ( 21 ): WkDay-1, WkDay-4, WkDay-24, WkDay-64, WkDay-106, WkDay-163, WkD
## FA anomalies ( 8 ): WkDay-24, WkDay-163, WkDay-205, WkDay-217, WkDay-274, WkDay-330,
## ICA anomalies ( 6 ): WkDay-81, WkDay-106, WkDay-175, WkDay-230, WkDay-232, WkDay-364
##
##
## =====
## Location: Loc15
## = = = = =
## PCA: retained 8 components
## PCA anomalies ( 20 ): WkDay-1, WkDay-5, WkDay-33, WkDay-66, WkDay-80, WkDay-83, WkDay
## FA anomalies ( 10 ): WkDay-1, WkDay-5, WkDay-33, WkDay-161, WkDay-232, WkDay-308, WkD
## ICA anomalies ( 8 ): WkDay-1, WkDay-161, WkDay-184, WkDay-214, WkDay-316, WkDay-317,
##
##
## =====
## Location: Loc16
## = = = = =
## PCA: retained 9 components
## PCA anomalies ( 29 ): WkDay-1, WkDay-9, WkDay-36, WkDay-45, WkDay-62, WkDay-77, WkDay
## FA anomalies ( 10 ): WkDay-1, WkDay-178, WkDay-230, WkDay-232, WkDay-260, WkDay-317,
## ICA anomalies ( 11 ): WkDay-9, WkDay-78, WkDay-191, WkDay-213, WkDay-222, WkDay-229,
##
##
## =====
## Location: Loc17
## = = = = =
## PCA: retained 8 components
## PCA anomalies ( 31 ): WkDay-1, WkDay-2, WkDay-6, WkDay-13, WkDay-46, WkDay-54, WkDay-
## FA anomalies ( 12 ): WkDay-1, WkDay-6, WkDay-13, WkDay-66, WkDay-80, WkDay-106, WkDay
## ICA anomalies ( 8 ): WkDay-2, WkDay-24, WkDay-54, WkDay-109, WkDay-212, WkDay-232, Wk
##
##
## =====
## Location: Loc18
## = = = = =
## PCA: retained 5 components
## PCA anomalies ( 15 ): WkDay-1, WkDay-26, WkDay-51, WkDay-93, WkDay-184, WkDay-199, Wk
## FA anomalies ( 0 ):
## ICA anomalies ( 5 ): WkDay-1, WkDay-239, WkDay-258, WkDay-259, WkDay-325
##
##

```

```

## =====
## Location: Loc19
## = = = = =
## PCA: retained 9 components
## PCA anomalies ( 21 ): WkDay-1, WkDay-14, WkDay-74, WkDay-76, WkDay-80, WkDay-114, WkD
## FA anomalies ( 10 ): WkDay-14, WkDay-80, WkDay-104, WkDay-114, WkDay-117, WkDay-176,
## ICA anomalies ( 13 ): WkDay-1, WkDay-24, WkDay-28, WkDay-74, WkDay-82, WkDay-137, WkD
##
##
## =====
## Location: Loc20
## = = = = =
## PCA: retained 6 components
## PCA anomalies ( 19 ): WkDay-1, WkDay-27, WkDay-29, WkDay-31, WkDay-49, WkDay-80, WkD
## FA anomalies ( 6 ): WkDay-1, WkDay-29, WkDay-31, WkDay-49, WkDay-354, WkDay-377
## ICA anomalies ( 4 ): WkDay-1, WkDay-205, WkDay-206, WkDay-212
##
##
## =====
## Location: Loc21
## = = = = =
## PCA: retained 5 components
## PCA anomalies ( 10 ): WkDay-83, WkDay-133, WkDay-188, WkDay-215, WkDay-223, WkDay-224
## FA anomalies ( 0 ):
## ICA anomalies ( 7 ): WkDay-1, WkDay-52, WkDay-232, WkDay-264, WkDay-327, WkDay-328, W
##
##
## =====
## Location: Loc22
## = = = = =
## PCA: retained 6 components
## PCA anomalies ( 13 ): WkDay-1, WkDay-11, WkDay-69, WkDay-70, WkDay-127, WkDay-140, Wk
## FA anomalies ( 10 ): WkDay-11, WkDay-69, WkDay-70, WkDay-140, WkDay-159, WkDay-221, W
## ICA anomalies ( 13 ): WkDay-1, WkDay-2, WkDay-69, WkDay-71, WkDay-127, WkDay-222, WkD
##
##
## =====
## Location: Loc23
## = = = = =
## PCA: retained 6 components
## PCA anomalies ( 20 ): WkDay-1, WkDay-33, WkDay-35, WkDay-47, WkDay-137, WkDay-156, Wk
## FA anomalies ( 16 ): WkDay-1, WkDay-33, WkDay-37, WkDay-38, WkDay-47, WkDay-137, WkD
## ICA anomalies ( 19 ): WkDay-1, WkDay-33, WkDay-37, WkDay-38, WkDay-137, WkDay-176, Wk
##
##

```

```
## =====
## Location: Loc24
## = = = = =
## PCA: retained 3 components
## PCA anomalies ( 5 ): WkDay-97, WkDay-162, WkDay-251, WkDay-278, WkDay-340
## FA anomalies ( 0 ):
## ICA anomalies ( 12 ): WkDay-57, WkDay-63, WkDay-100, WkDay-103, WkDay-162, WkDay-239,
##
##
## =====
## Location: Loc25
## = = = = =
## PCA: retained 7 components
## PCA anomalies ( 22 ): WkDay-1, WkDay-31, WkDay-33, WkDay-35, WkDay-36, WkDay-37, WkDay-38, WkDay-39, WkDay-40, WkDay-41, WkDay-42, WkDay-43, WkDay-44, WkDay-45, WkDay-46, WkDay-47, WkDay-48, WkDay-49, WkDay-50, WkDay-51, WkDay-52, WkDay-53, WkDay-54, WkDay-55
## FA anomalies ( 6 ): WkDay-95, WkDay-102, WkDay-251, WkDay-257, WkDay-295, WkDay-384
## ICA anomalies ( 19 ): WkDay-35, WkDay-36, WkDay-37, WkDay-47, WkDay-52, WkDay-63, WkDay-64, WkDay-65, WkDay-66, WkDay-67, WkDay-68, WkDay-69, WkDay-70, WkDay-71, WkDay-72, WkDay-73, WkDay-74, WkDay-75, WkDay-76
##
##
## =====
## Location: Loc26
## = = = = =
## PCA: retained 2 components
## PCA anomalies ( 6 ): WkDay-1, WkDay-33, WkDay-226, WkDay-227, WkDay-235, WkDay-239
## FA anomalies ( 0 ):
## ICA anomalies ( 7 ): WkDay-18, WkDay-221, WkDay-222, WkDay-226, WkDay-231, WkDay-234, WkDay-235
```

Plot PCA Loadings

```
plot_pca_loadings <- function(res, loc_name, n_comp = 3) {  
  loadings <- res$pca_loadings[, 1:min(n_comp, ncol(res$pca_loadings)), drop = FALSE]  
  
  # Convert time intervals to hours (288 intervals = 24 hours)  
  time_hours <- seq(0, 24, length.out = nrow(loadings))  
  
  df <- data.frame(Time = time_hours, loadings)  
  colnames(df) <- c("Time", paste0("PC", 1:ncol(loadings)))  
  
  df_long <- tidyr::pivot_longer(df, cols = -Time,  
                                names_to = "Component",  
                                values_to = "Loading")  
  
  ggplot(df_long, aes(x = Time, y = Loading, color = Component)) +  
    geom_line(size = 1) +
```



```

    facet_wrap(~Component, ncol = 1, scales = "free_y") +
    labs(title = paste("PCA Loadings -", loc_name),
         x = "Hour of Day",
         y = "Loading") +
    theme_minimal() +
    scale_x_continuous(breaks = seq(0, 24, 4))
}

for (loc_name in names(location_results)) {
  print(plot_pca_loadings(location_results[[loc_name]], loc_name))
}

```

View Anomalies with Boxplot

```

plot_pca_score_boxplot <- function(res, loc_name, include_fa = FALSE) {
  scores <- res$pca_scores
  days <- rownames(scores)

  df_long <- data.frame(
    Day = rep(days, ncol(scores)),
    PC = rep(colnames(scores), each = nrow(scores)),
    Score = c(scores)
  )

  # Identify anomalies
  df_long$IsPCAAnomaly <- ifelse(df_long$Day %in% res$pca_anom_days, "PCA Anomaly", "Normal")

  # OPTIONALLY include FA anomalies
  if (include_fa && !is.null(res$fa_scores)) {
    df_long$IsFAAnomaly <- ifelse(df_long$Day %in% res$fa_anom_days, "FA Anomaly", "Normal")
    df_long$AnomalyType <- ifelse(df_long$IsPCAAnomaly == "PCA Anomaly",
                                  "PCA Anomaly",
                                  ifelse(df_long$IsFAAnomaly == "FA Anomaly", "FA Anomaly", "Normal"))
  } else {
    df_long$AnomalyType <- df_long$IsPCAAnomaly
  }

  ggplot(df_long, aes(x = PC, y = Score)) +
    geom_boxplot(outlier.shape = NA) + # Remove default black dots
    geom_jitter(aes(color = AnomalyType), alpha = 0.6, width = 0.15, size = 2) +
    scale_color_manual(values = c("Normal" = "gray",
                                   "PCA Anomaly" = "red",
                                   "FA Anomaly" = "red"))
}

```

```

                                "FA Anomaly" = "blue")) +
  labs(title = paste("PC Score Boxplots with Anomalies -", loc_name),
        x = "Principal Component",
        y = "Score",
        color = "Detection Method") +
  theme_minimal() +
  theme(legend.position = "bottom")
}

for (loc_name in names(location_results)) {
  print(plot_pca_score_boxplot(location_results[[loc_name]], loc_name, include_fa=TRUE))
}

```

Anomalies overlay boxplot with labels

```

overlay <- function(res, loc_name, include_fa = FALSE) {
  scores <- res$pca_scores
  days <- rownames(scores)

  df_long <- data.frame(
    Day = rep(days, ncol(scores)),
    PC = rep(colnames(scores), each = nrow(scores)),
    Score = c(scores)
  )

  # Correct anomaly identification
  if (include_fa && !is.null(res$fa_scores)) {
    df_long$AnomalyType <- case_when(
      df_long$Day %in% intersect(res$pca_anom_days, res$fa_anom_days) ~ "Both PCA & FA",
      df_long$Day %in% setdiff(res$pca_anom_days, res$fa_anom_days) ~ "PCA Anomaly",
      df_long$Day %in% setdiff(res$fa_anom_days, res$pca_anom_days) ~ "FA Anomaly",
      TRUE ~ "Normal"
    )
  } else {
    df_long$AnomalyType <- ifelse(df_long$Day %in% res$pca_anom_days, "PCA Anomaly", "Normal")
  }

  ggplot(df_long, aes(x = PC, y = Score)) +
    geom_boxplot(outlier.shape = NA) +

    # Normal points first (light and small)
    geom_jitter(data = subset(df_long, AnomalyType == "Normal"),

```

```

        color = "gray70", alpha = 0.3, width = 0.15, size = 1) +

# PCA-only anomalies
geom_point(data = subset(df_long, AnomalyType == "PCA Anomaly"),
           aes(color = "PCA Anomaly"), size = 1.0, alpha = 0.9) +

# FA-only anomalies
geom_point(data = subset(df_long, AnomalyType == "FA Anomaly"),
           aes(color = "FA Anomaly"), size = 1.0, alpha = 0.9) +

# Both PCA & FA
geom_point(data = subset(df_long, AnomalyType == "Both PCA & FA"),
           aes(color = "Both PCA & FA"), size = 1.3, alpha = 1, shape = 8) +

scale_color_manual(values = c(
  "PCA Anomaly" = "red",
  "FA Anomaly" = "blue",
  "Both PCA & FA" = "purple"
)) +
labs(title = paste("Anomalies Highlighted Clearly -", loc_name),
     x = "Principal Component", y = "Score", color = "Anomaly Source") +
theme_minimal() +
theme(legend.position = "bottom")
}

for (loc_name in names(location_results)) {
  print(overlay(location_results[[loc_name]], loc_name, include_fa=TRUE))
}

```

visualize ica loadings

```

## ICA Source Loadings Plot
plot_ica_loadings <- function(res, loc_name, max_plot_comp = 4) {
  if (is.null(res$ica) || is.null(res$ica$A) || ncol(res$ica$A) == 0) {
    message("ICA Loadings (A matrix) not available for ", loc_name)
    return(NULL)
  }

  # ICA Loadings are stored in the 'A' matrix
  loadings <- res$ica$A
  n_comp <- ncol(loadings)

```

```

# Limit components being plotted for stability
n_comp_to_plot <- min(n_comp, max_plot_comp)
loadings <- loadings[, 1:n_comp_to_plot, drop = FALSE]

# Convert time intervals to hours
time_hours <- seq(0, 24, length.out = nrow(loadings))

df <- data.frame(Time = time_hours, loadings)
colnames(df) <- c("Time", paste0("Source", 1:n_comp_to_plot))

df_long <- tidyr::pivot_longer(df, cols = -Time,
                              names_to = "Source",
                              values_to = "Loading")

ggplot(df_long, aes(x = Time, y = Loading, color = Source)) +
  geom_line(size = 1) +
  facet_wrap(~Source, ncol = 1, scales = "free_y") +
  labs(title = paste("ICA Source Loadings (A Matrix) -", loc_name),
       x = "Hour of Day",
       y = "Loading/Mixing Weight") +
  theme_minimal() +
  scale_x_continuous(breaks = seq(0, 24, 4))
}

# Loop over locations to generate the plots
for (loc_name in names(location_results)) {
  # Wrap in tryCatch for final safety
  tryCatch({
    print(plot_ica_loadings(location_results[[loc_name]], loc_name))
  }, error = function(e) {
    warning(paste("Failed to plot ICA Loadings for", loc_name, ":", e$message))
  })
}

```

Better: Plot Anomalous Days vs Normal Days

```

anomaly_timeline <- function(res, loc_name) {
  all_days <- rownames(res$pca_scores)
  df_timeline <- data.frame(DayName = all_days, DayIndex = 1:length(all_days))

  df_timeline <- df_timeline %>%
    mutate(AnomalyType = case_when(

```

```

DayName %in% intersect(intersect(res$pca_anom_days, res$fa_anom_days), res$ica_anom_days)
DayName %in% setdiff(intersect(res$pca_anom_days, res$fa_anom_days), res$ica_anom_days)
DayName %in% setdiff(intersect(res$pca_anom_days, res$ica_anom_days), res$fa_anom_days)
DayName %in% setdiff(intersect(res$fa_anom_days, res$ica_anom_days), res$pca_anom_days)
DayName %in% setdiff(setdiff(res$pca_anom_days, res$fa_anom_days), res$ica_anom_days)
DayName %in% setdiff(setdiff(res$fa_anom_days, res$pca_anom_days), res$ica_anom_days)
DayName %in% setdiff(setdiff(res$ica_anom_days, res$pca_anom_days), res$fa_anom_days)
TRUE ~ "Normal"
))

df_anomalies <- df_timeline %>% filter(AnomalyType != "Normal")

if (nrow(df_anomalies) == 0) {
  message("No anomalies to plot for ", loc_name)
  return(NULL)
}

ggplot(df_anomalies, aes(x = DayIndex, y = 1)) +

  # Clean horizontal baseline
  geom_hline(yintercept = 1, color = "gray85", linewidth = 0.3) +

  # anomaly points only
  geom_point(aes(color = AnomalyType, shape = AnomalyType), size = 3) +

  scale_color_manual(values = c(
    "PCA Only" = "red",
    "FA Only" = "blue",
    "ICA Only" = "green",
    "PCA & FA" = "purple",
    "PCA & ICA" = "#FF7F50",
    "FA & ICA" = "#00CED1",
    "All Three" = "black"
  )) +

  scale_shape_manual(values = c(
    "PCA Only" = 19,
    "FA Only" = 17,
    "ICA Only" = 15,
    "PCA & FA" = 8,
    "PCA & ICA" = 10,
    "FA & ICA" = 13,
    "All Three" = 4
  )) +

  labs(title = paste("Anomaly Temporal Sequence -", loc_name),

```

```

    x = "Day Index (Time Sequence)",
    y = "",
    color = "Detection Source",
    shape = "Detection Source") +
theme_minimal() +
theme(axis.text.y = element_blank(),
      axis.ticks.y = element_blank(),
      panel.grid.major.y = element_blank(),
      panel.grid.minor.y = element_blank(),
      legend.position = "bottom")
}

# Plot for all locations
for (loc_name in names(location_results)) {
  print(anomaly_timeline(location_results[[loc_name]], loc_name))
}

```

View Summary of All Locations

```

# Create summary table
summary_df <- data.frame(
  Location = names(location_results),

  # Total unique anomalies across all methods
  Total_Anomalies = sapply(location_results, \(x)
    length(union(union(x$pca_anom_days, x$fa_anom_days), x$ica_anom_days))),

  # All three methods detected
  All_Three = sapply(location_results, \(x)
    length(intersect(intersect(x$pca_anom_days, x$fa_anom_days), x$ica_anom_days))),

  # Two methods detected
  PCA_FA = sapply(location_results, \(x)
    length(setdiff(intersect(x$pca_anom_days, x$fa_anom_days), x$ica_anom_days))),

  PCA_ICA = sapply(location_results, \(x)
    length(setdiff(intersect(x$pca_anom_days, x$ica_anom_days), x$fa_anom_days))),

  FA_ICA = sapply(location_results, \(x)
    length(setdiff(intersect(x$fa_anom_days, x$ica_anom_days), x$pca_anom_days))),

  # Single method only

```

```

PCA_Only = sapply(location_results, \(x)
  length(setdiff(setdiff(x$pca_anom_days, x$fa_anom_days), x$ica_anom_days))),

FA_Only = sapply(location_results, \(x)
  length(setdiff(setdiff(x$fa_anom_days, x$pca_anom_days), x$ica_anom_days))),

ICA_Only = sapply(location_results, \(x)
  length(setdiff(setdiff(x$ica_anom_days, x$pca_anom_days), x$fa_anom_days))),

FA_Success = sapply(location_results, \(x) !is.null(x$fa_model)),
ICA_Success = sapply(location_results, \(x) !is.null(x$ica_scores))
)

df_long_summary <- summary_df %>%
  pivot_longer(cols = -Location,
    names_to = "AnomalyType",
    values_to = "Count") %>%
  filter(Count > 0)

print(summary_df)

```

##	Location	Total_Anomalies	All_Three	PCA_FA	PCA_ICA	FA_ICA	PCA_Only	FA_Only
## Loc1	Loc1	39	0	0	8	0	22	0
## Loc2	Loc2	27	0	8	9	0	3	2
## Loc3	Loc3	31	0	0	10	0	14	0
## Loc4	Loc4	27	3	7	1	0	9	1
## Loc5	Loc5	23	3	1	1	0	13	1
## Loc6	Loc6	6	3	1	0	0	2	0
## Loc7	Loc7	15	0	0	4	0	0	0
## Loc8	Loc8	25	0	9	5	0	9	1
## Loc9	Loc9	17	0	0	9	0	5	0
## Loc10	Loc10	17	0	0	3	0	10	0
## Loc11	Loc11	28	3	7	2	0	12	0
## Loc12	Loc12	29	8	4	5	2	6	0
## Loc13	Loc13	32	5	1	3	0	11	3
## Loc14	Loc14	26	0	7	2	0	12	1
## Loc15	Loc15	24	4	4	2	0	10	2
## Loc16	Loc16	39	0	6	5	0	18	4
## Loc17	Loc17	36	0	11	4	0	16	1
## Loc18	Loc18	17	0	0	3	0	12	0
## Loc19	Loc19	30	1	8	4	0	8	1
## Loc20	Loc20	21	1	4	2	0	12	1

## Loc21	Loc21	14	0	0	3	0	7	0
## Loc22	Loc22	22	1	8	4	0	0	1
## Loc23	Loc23	26	9	2	5	4	4	1
## Loc24	Loc24	15	0	0	2	0	3	0
## Loc25	Loc25	34	4	2	3	0	13	0
## Loc26	Loc26	12	0	0	1	0	5	0
##	ICA_Only	FA_Success	ICA_Success					
## Loc1	9	FALSE	TRUE					
## Loc2	5	TRUE	TRUE					
## Loc3	7	FALSE	TRUE					
## Loc4	6	TRUE	TRUE					
## Loc5	4	TRUE	TRUE					
## Loc6	0	TRUE	TRUE					
## Loc7	11	FALSE	TRUE					
## Loc8	1	TRUE	TRUE					
## Loc9	3	FALSE	TRUE					
## Loc10	4	FALSE	TRUE					
## Loc11	4	TRUE	TRUE					
## Loc12	4	TRUE	TRUE					
## Loc13	9	TRUE	TRUE					
## Loc14	4	TRUE	TRUE					
## Loc15	2	TRUE	TRUE					
## Loc16	6	TRUE	TRUE					
## Loc17	4	TRUE	TRUE					
## Loc18	2	FALSE	TRUE					
## Loc19	8	TRUE	TRUE					
## Loc20	1	TRUE	TRUE					
## Loc21	4	FALSE	TRUE					
## Loc22	8	TRUE	TRUE					
## Loc23	1	TRUE	TRUE					
## Loc24	10	FALSE	TRUE					
## Loc25	12	TRUE	TRUE					
## Loc26	6	FALSE	TRUE					

Stacked bar chart of anomaly counts

```
df_long_summary <- summary_df %>%
  pivot_longer(cols = c(PCA_Only, FA_Only, ICA_Only,
                        PCA_FA, PCA_ICA, FA_ICA, All_Three),
               names_to = "AnomalyType",
               values_to = "Count") %>%
  filter(Count > 0)
```



```

fill_colors <- c(
  "PCA_Only" = "red",
  "FA_Only" = "blue",
  "ICA_Only" = "green",
  "PCA_FA" = "purple",
  "PCA_ICA" = "#FF7F50",
  "FA_ICA" = "#00CED1",
  "All_Three" = "black"
)

ggplot(df_long_summary, aes(x = Location, y = Count, fill = AnomalyType)) +
  geom_bar(stat = "identity") +
  scale_fill_manual(values = fill_colors,
    labels = c(
      "PCA_Only" = "PCA Only",
      "FA_Only" = "FA Only",
      "ICA_Only" = "ICA Only",
      "PCA_FA" = "PCA & FA",
      "PCA_ICA" = "PCA & ICA",
      "FA_ICA" = "FA & ICA",
      "All_Three" = "All 3 Methods"
    )) +
  labs(title = "Breakdown of Anomaly Types by Location",
    x = "Location",
    y = "Number of Anomalous Days",
    fill = "Anomaly Source") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))

```

normalized

```

# Compute proportion (normalized values)
df_long_summary <- df_long_summary %>%
  group_by(Location) %>%
  mutate(Proportion = Count / sum(Count)) %>%
  ungroup()

# Define colors for each category
fill_colors <- c(
  "PCA_Only" = "red",
  "FA_Only" = "blue",
  "ICA_Only" = "green",

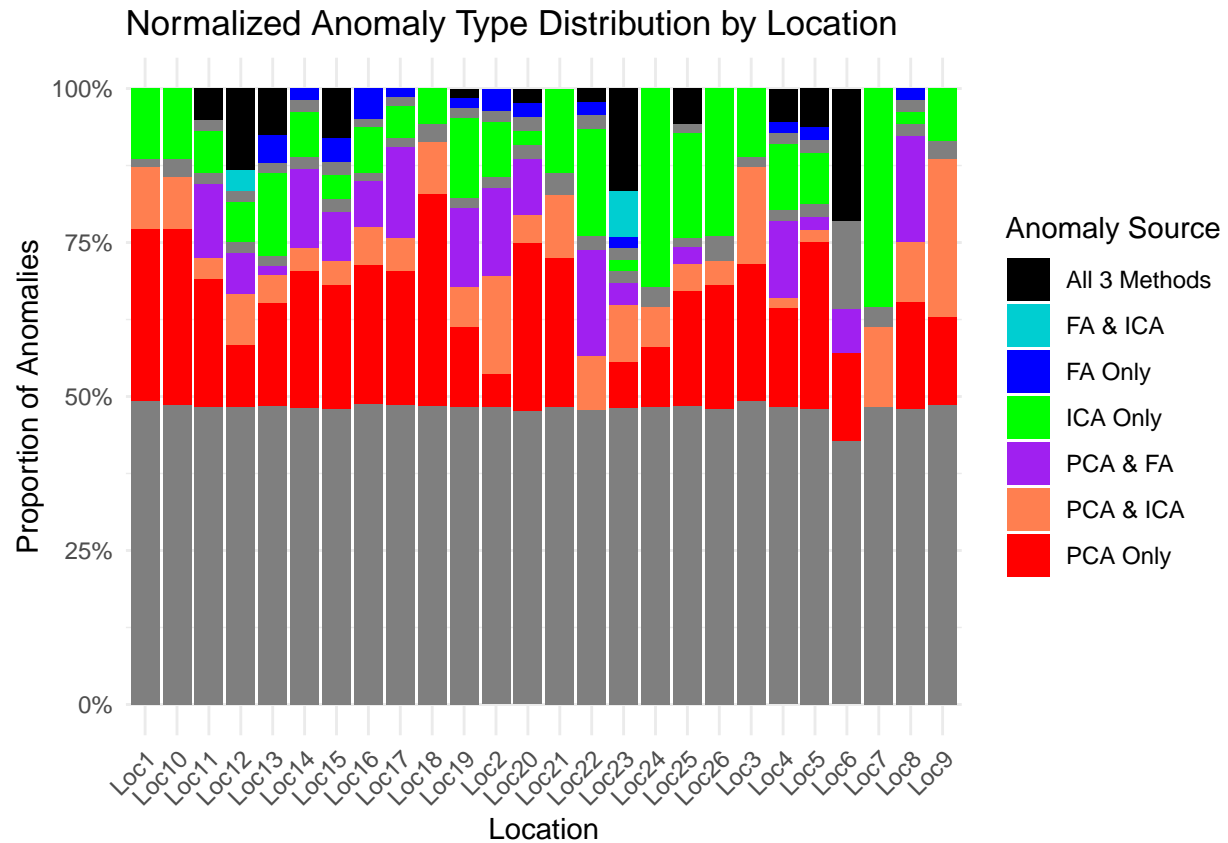
```

```

"PCA_FA" = "purple",
"PCA_ICA" = "#FF7F50",
"FA_ICA" = "#00CED1",
"All_Three" = "black"
)

# Plot normalized stacked bar chart
ggplot(df_long_summary, aes(x = Location, y = Proportion, fill = AnomalyType)) +
  geom_bar(stat = "identity") +
  scale_fill_manual(values = fill_colors,
                    labels = c(
                      "PCA_Only" = "PCA Only",
                      "FA_Only" = "FA Only",
                      "ICA_Only" = "ICA Only",
                      "PCA_FA" = "PCA & FA",
                      "PCA_ICA" = "PCA & ICA",
                      "FA_ICA" = "FA & ICA",
                      "All_Three" = "All 3 Methods"
                    )) +
  theme_minimal() +
  labs(title = "Normalized Anomaly Type Distribution by Location",
       x = "Location",
       y = "Proportion of Anomalies",
       fill = "Anomaly Source") +
  scale_y_continuous(labels = scales::percent_format()) +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))

```



component loadings heatmap

```
# R Code for Component Loadings Heatmap (Example: Loc1):

loc_name <- names(location_results)[1]
res <- location_results[[loc_name]]
n_comp_to_plot <- min(3, ncol(res$pca_loadings))
loadings <- res$pca_loadings[, 1:n_comp_to_plot, drop = FALSE]

# Convert time intervals to index (1 to 288) and then approximate Hour
time_intervals <- 1:nrow(loadings)

df_loadings <- data.frame(TimeIndex = time_intervals, loadings)
colnames(df_loadings) <- c("TimeIndex", paste0("PC", 1:n_comp_to_plot))

df_long_loadings <- df_loadings %>%
  pivot_longer(cols = -TimeIndex,
               names_to = "Component",
               values_to = "Loading")

# Map TimeIndex to approximate Hour for axis readability (288 intervals = 24 hours)
df_long_loadings$Hour <- (df_long_loadings$TimeIndex - 1) * (24 / 288)
```

```

ggplot(df_long_loadings, aes(x = Hour, y = Component, fill = Loading)) +
  geom_tile(color = "white", linewidth = 0.5) +
  scale_fill_gradient2(low = "blue", mid = "white", high = "red", midpoint = 0,
    name = "Loading Value") +
  labs(title = paste("PCA Component Loadings Heatmap -", loc_name),
    x = "Hour of Day",
    y = "Principal Component") +
  scale_x_continuous(breaks = seq(0, 24, 4)) +
  theme_minimal() +
  theme(legend.position = "bottom",
    axis.text.y = element_text(face = "bold"))

```

Anomaly timeline

```

anomaly_timeline <- function(res, loc_name) {
  all_days <- rownames(res$pca_scores)
  df_timeline <- data.frame(DayName = all_days, DayIndex = 1:length(all_days))

  # Multi-method anomaly labeling
  df_timeline <- df_timeline %>%
    mutate(AnomalyType = case_when(
      DayName %in% intersect(intersect(res$pca_anom_days, res$fa_anom_days), res$ica_anom_days) ~ "Anomaly",
      DayName %in% setdiff(intersect(res$pca_anom_days, res$fa_anom_days), res$ica_anom_days) ~ "Anomaly",
      DayName %in% setdiff(intersect(res$pca_anom_days, res$ica_anom_days), res$fa_anom_days) ~ "Anomaly",
      DayName %in% setdiff(intersect(res$fa_anom_days, res$ica_anom_days), res$pca_anom_days) ~ "Anomaly",
      DayName %in% setdiff(setdiff(res$pca_anom_days, res$fa_anom_days), res$ica_anom_days) ~ "Anomaly",
      DayName %in% setdiff(setdiff(res$fa_anom_days, res$pca_anom_days), res$ica_anom_days) ~ "Anomaly",
      DayName %in% setdiff(setdiff(res$ica_anom_days, res$pca_anom_days), res$fa_anom_days) ~ "Anomaly",
      TRUE ~ "Normal"
    ))

  df_anomalies <- df_timeline %>% filter(AnomalyType != "Normal")

  if (nrow(df_anomalies) == 0) {
    message("No anomalies to plot for ", loc_name)
    return(NULL)
  }

  ggplot(df_timeline, aes(x = DayIndex, y = 1)) +
    # Normal timeline background
    # geom_segment(data = subset(df_timeline, AnomalyType == "Normal"),
    # aes(x = DayIndex, xend = DayIndex, y = 0.95, yend = 1.05),

```

```

#           color = "gray80", linewidth = 0.15) +

# Highlight anomalies
geom_point(data = df_anomalies,
           aes(color = AnomalyType, shape = AnomalyType),
           size = 3) +

scale_color_manual(values = c(
  "PCA Only" = "red",
  "FA Only" = "blue",
  "ICA Only" = "green",
  "PCA & FA" = "purple",
  "PCA & ICA" = "#FF7F50",
  "FA & ICA" = "#00CED1",
  "All Three" = "black"
)) +
scale_shape_manual(values = c(
  "PCA Only" = 19,      # circle
  "FA Only" = 17,      # triangle
  "ICA Only" = 15,     # square
  "PCA & FA" = 8,       # star
  "PCA & ICA" = 10,    # diamond
  "FA & ICA" = 13,     # diamond plus
  "All Three" = 4      # X
)) +

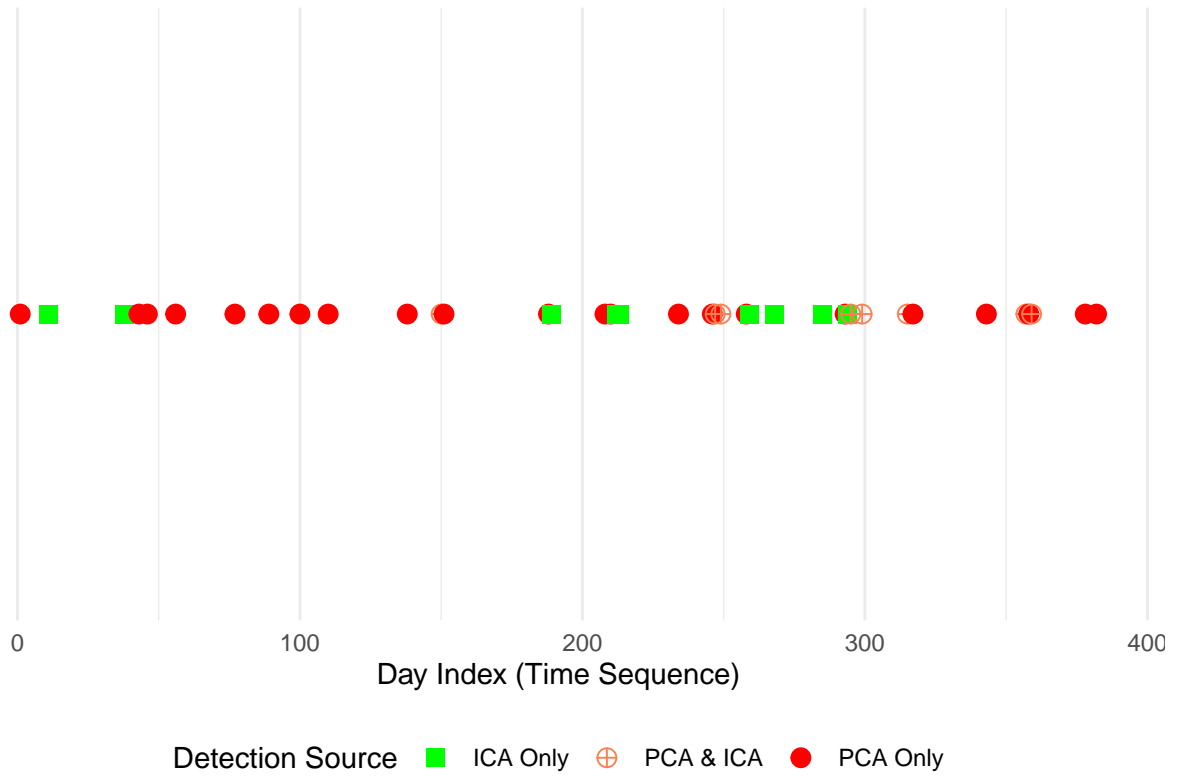
labs(title = paste("Anomaly Temporal Sequence -", loc_name),
     x = "Day Index (Time Sequence)",
     y = "",
     color = "Detection Source",
     shape = "Detection Source") +

theme_minimal() +
theme(axis.text.y = element_blank(),
      axis.ticks.y = element_blank(),
      panel.grid.major.y = element_blank(),
      panel.grid.minor.y = element_blank(),
      legend.position = "bottom")
}

# Loop over locations to generate the plots
for (loc_name in names(location_results)) {
  print(anomaly_timeline(location_results[[loc_name]], loc_name))
}

```

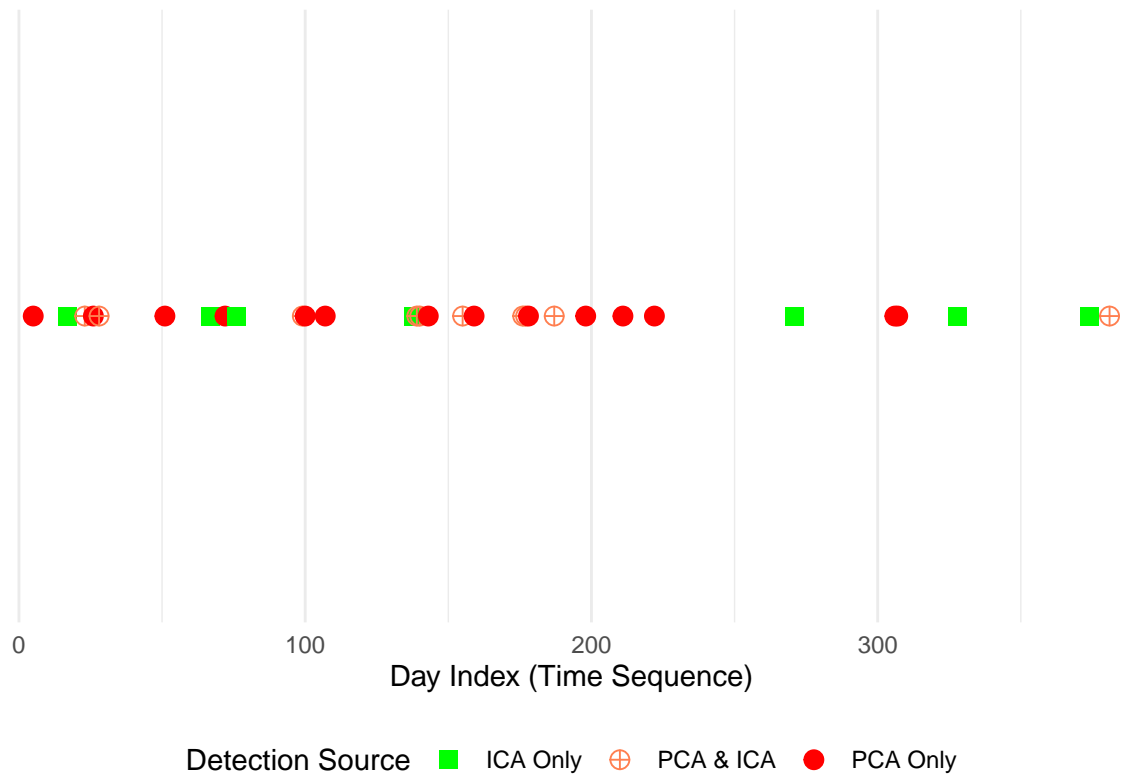
Anomaly Temporal Sequence – Loc1



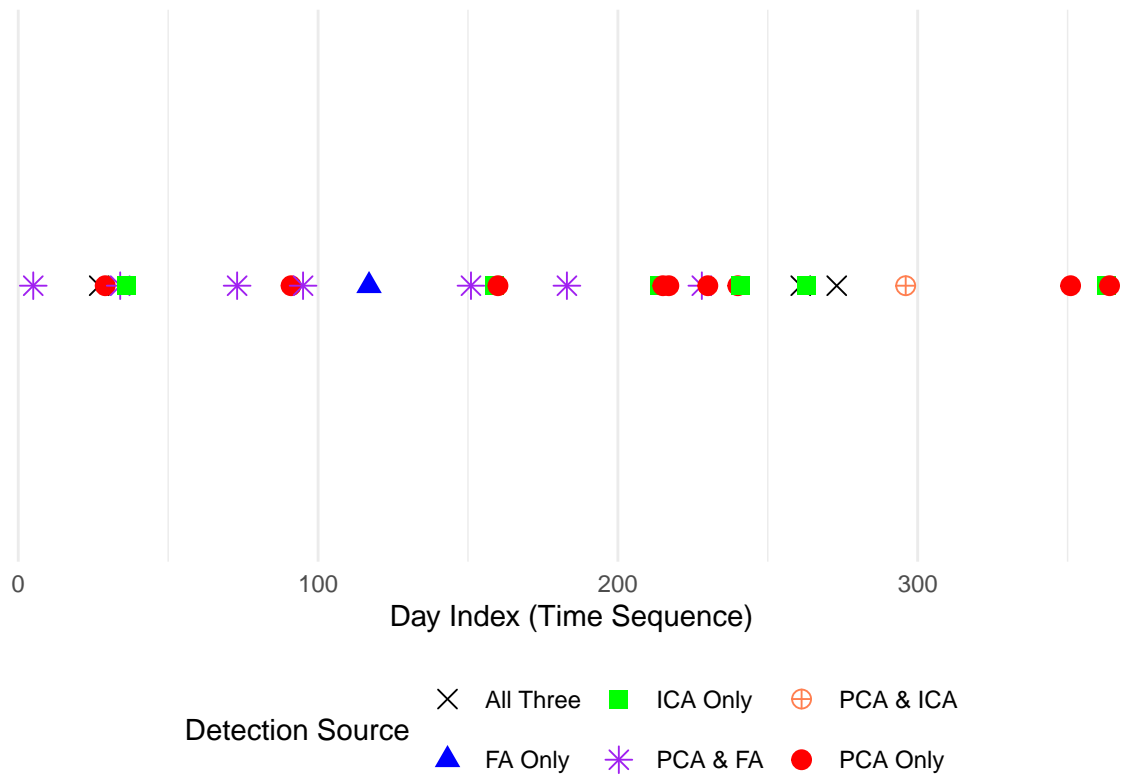
Anomaly Temporal Sequence – Loc2



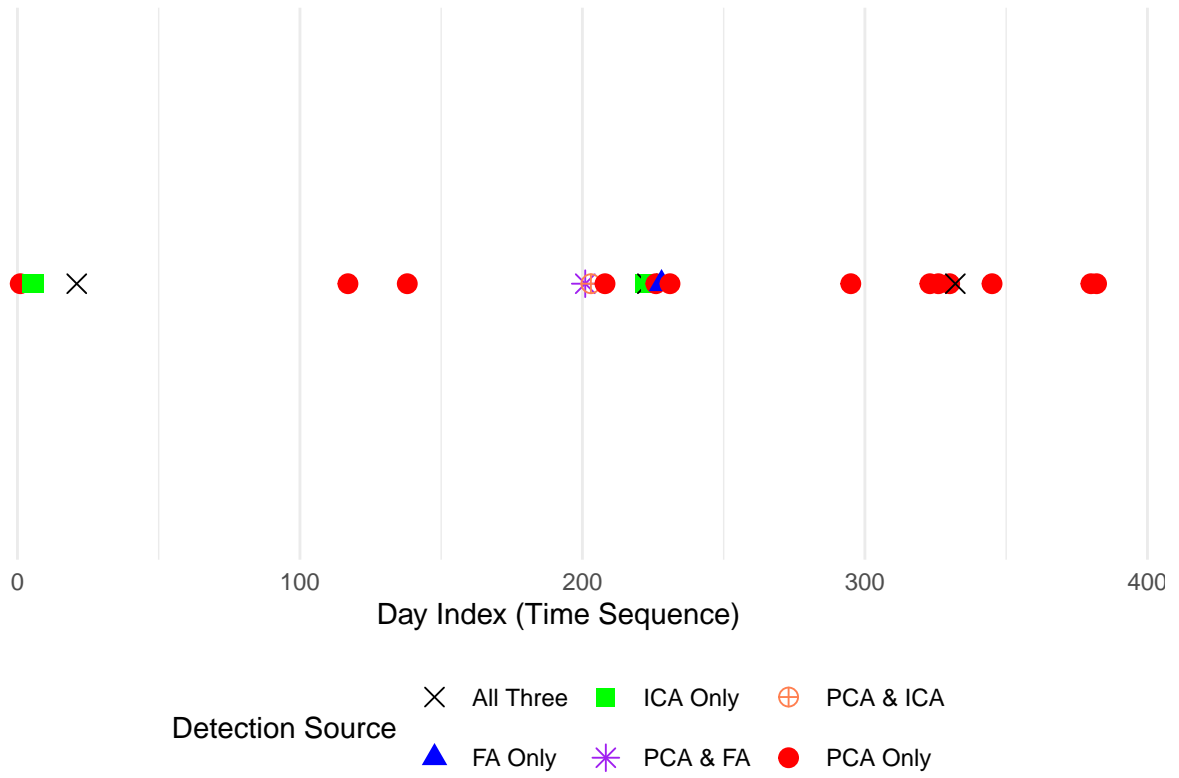
Anomaly Temporal Sequence – Loc3



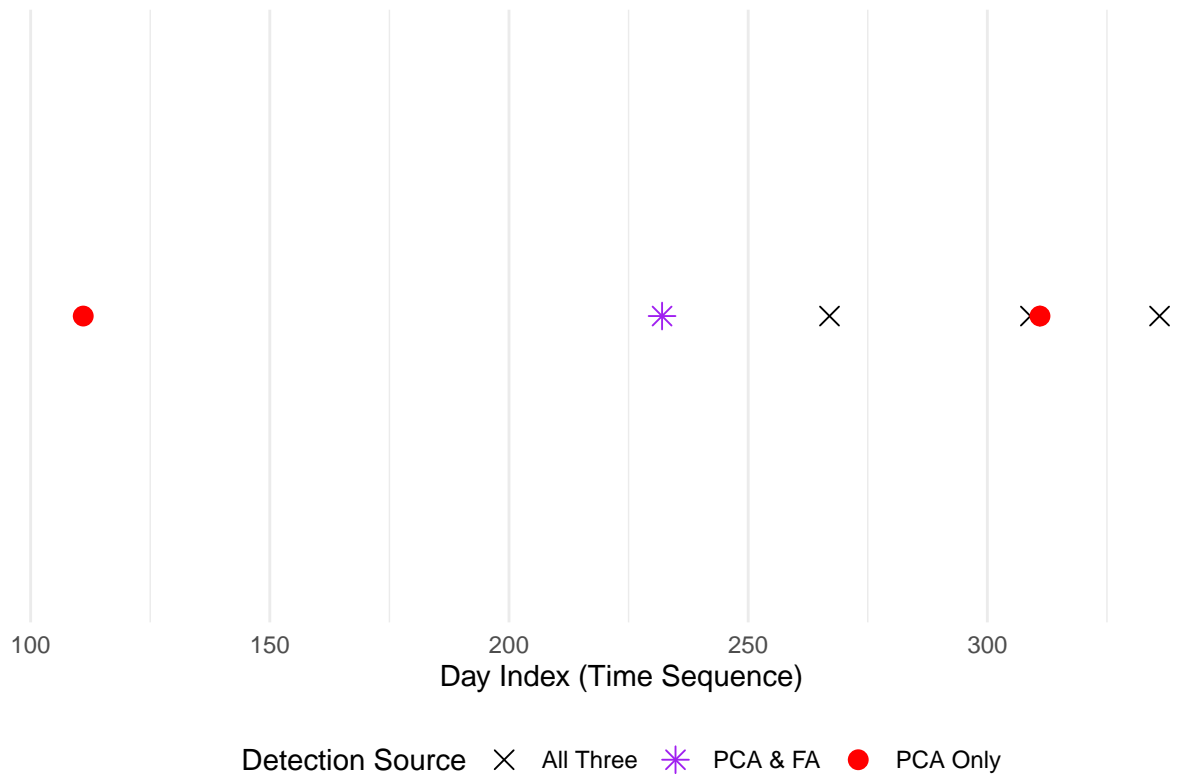
Anomaly Temporal Sequence – Loc4



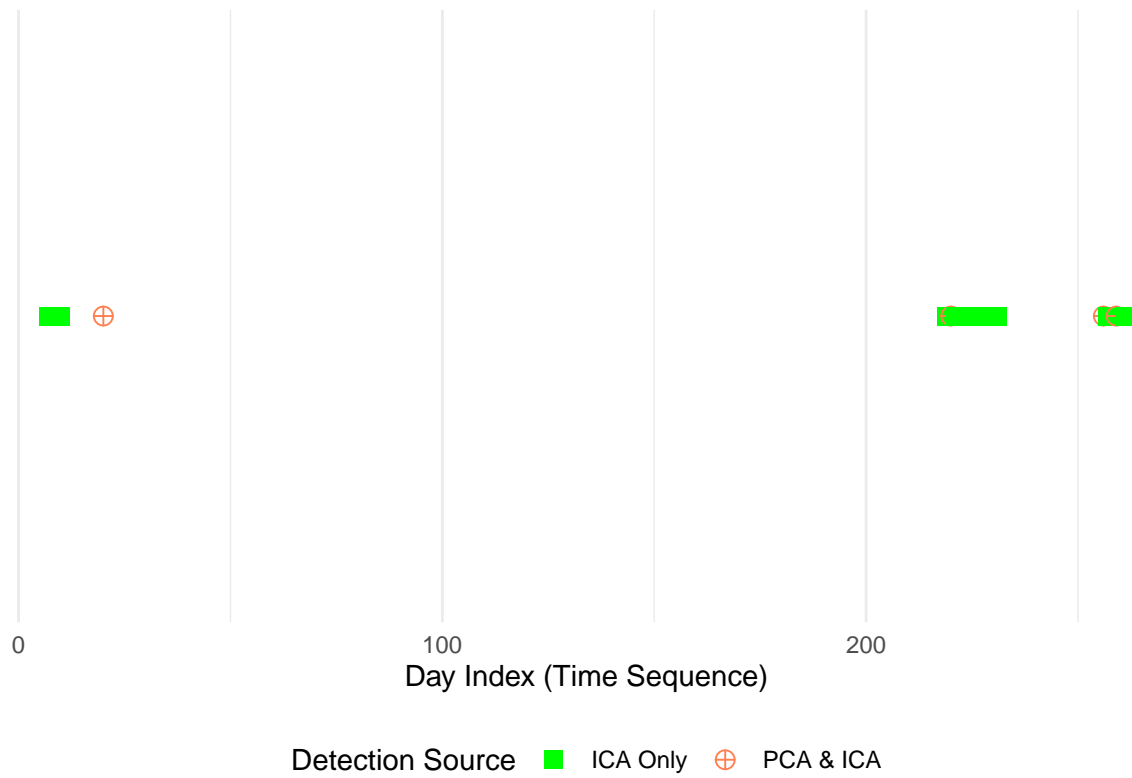
Anomaly Temporal Sequence – Loc5



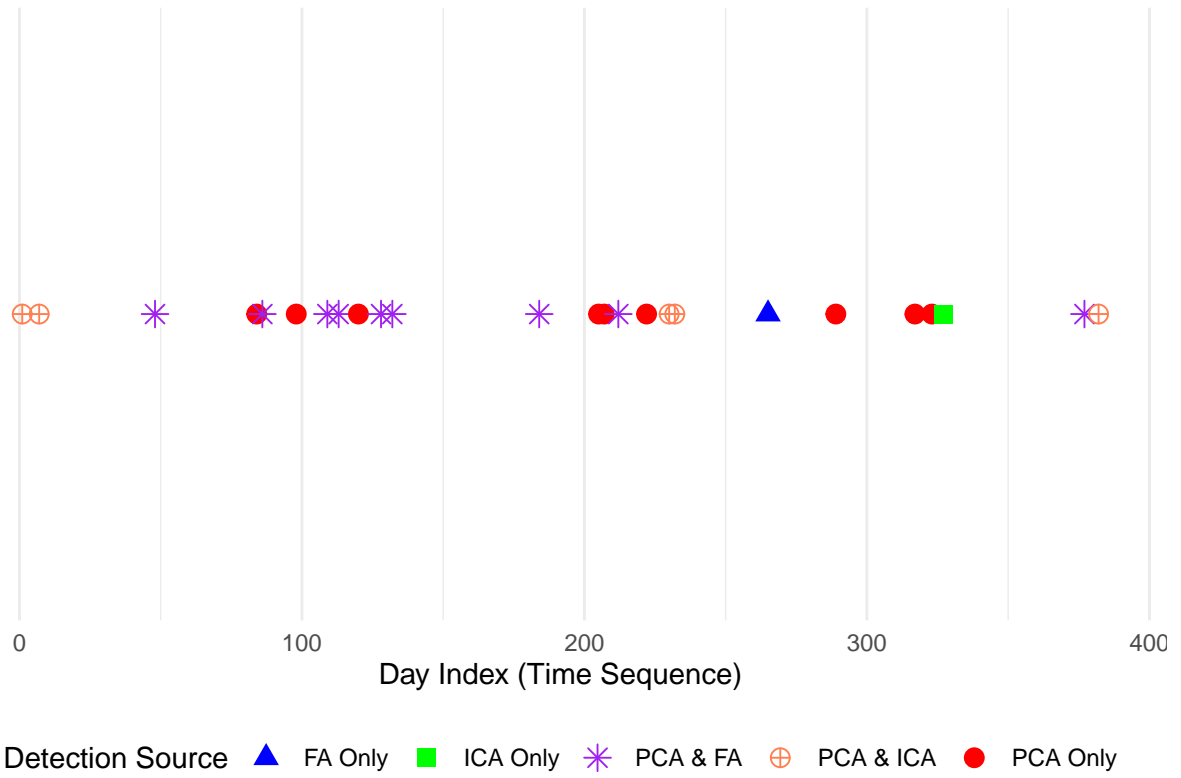
Anomaly Temporal Sequence – Loc6



Anomaly Temporal Sequence – Loc7



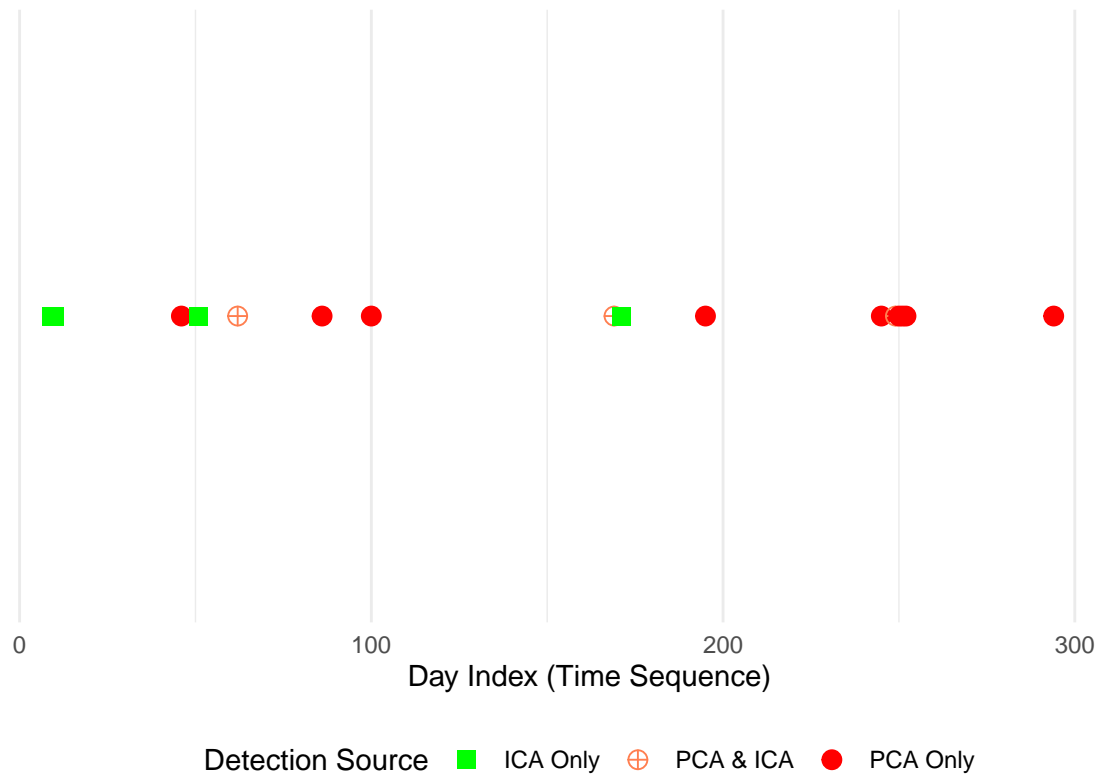
Anomaly Temporal Sequence – Loc8



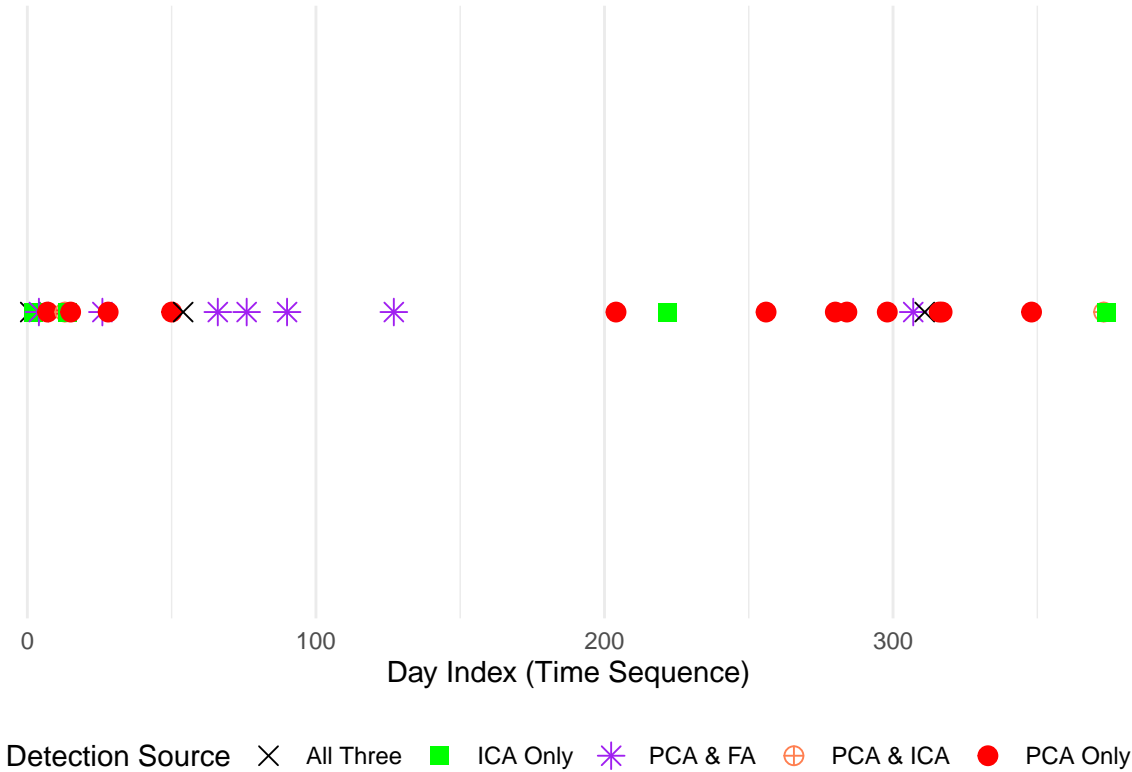
Anomaly Temporal Sequence – Loc9



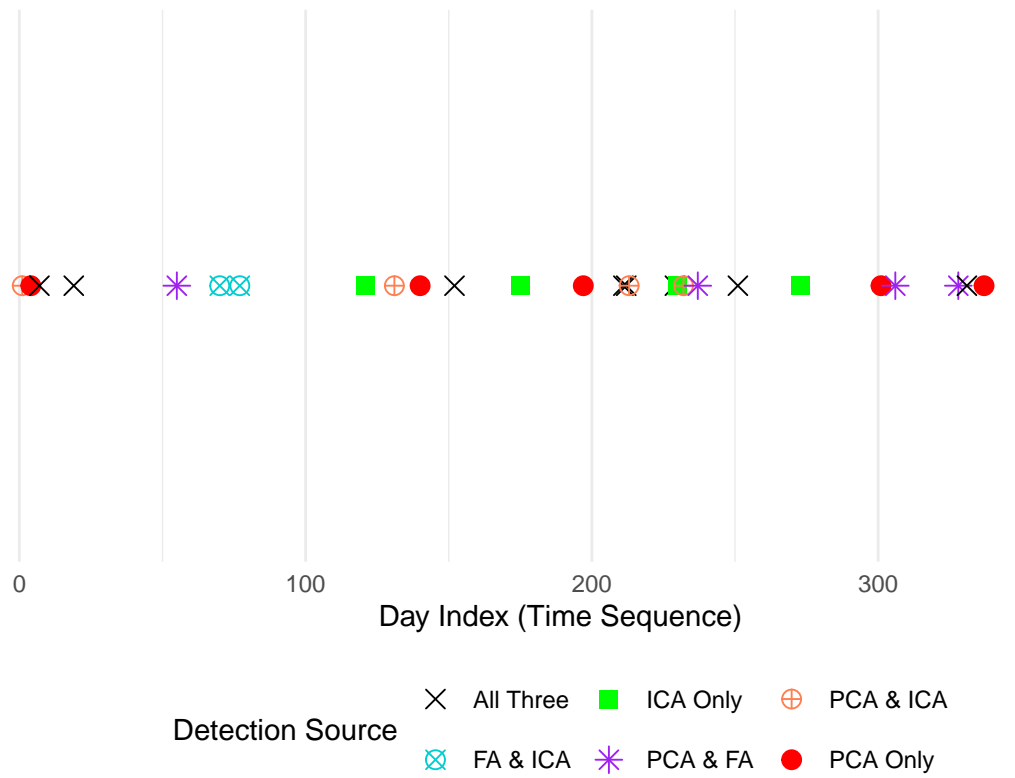
Anomaly Temporal Sequence – Loc10



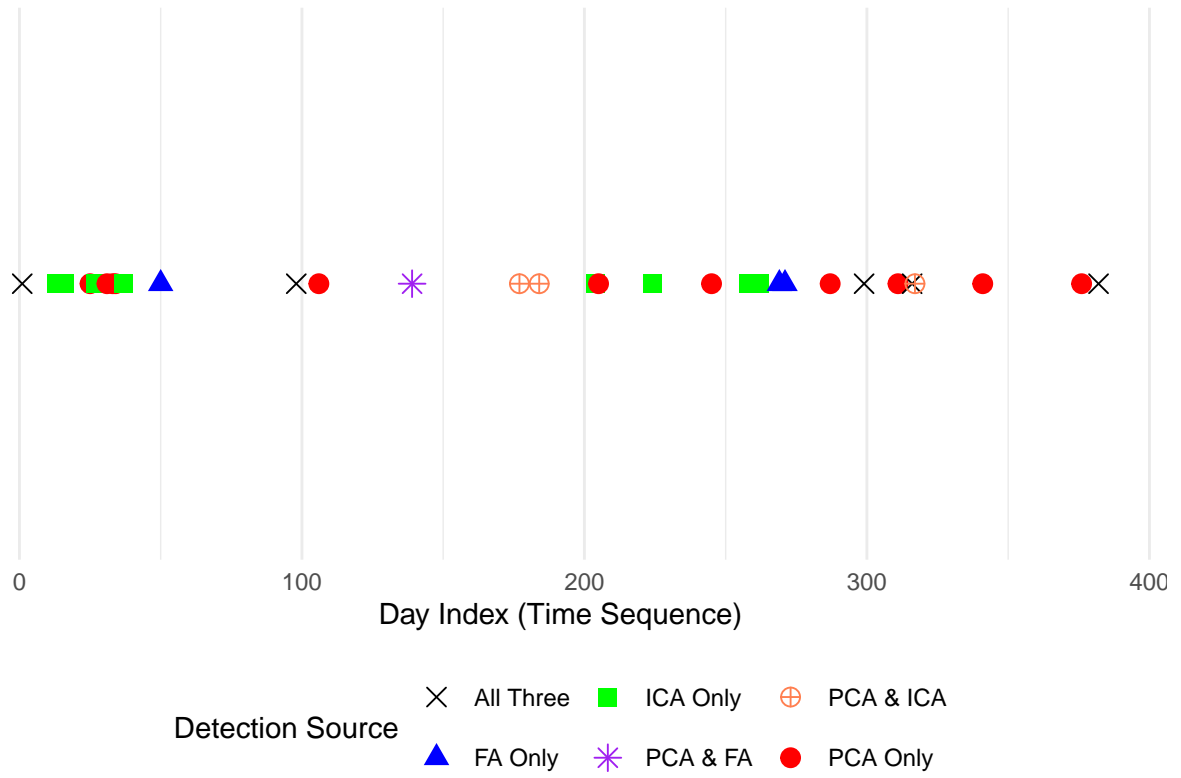
Anomaly Temporal Sequence – Loc11



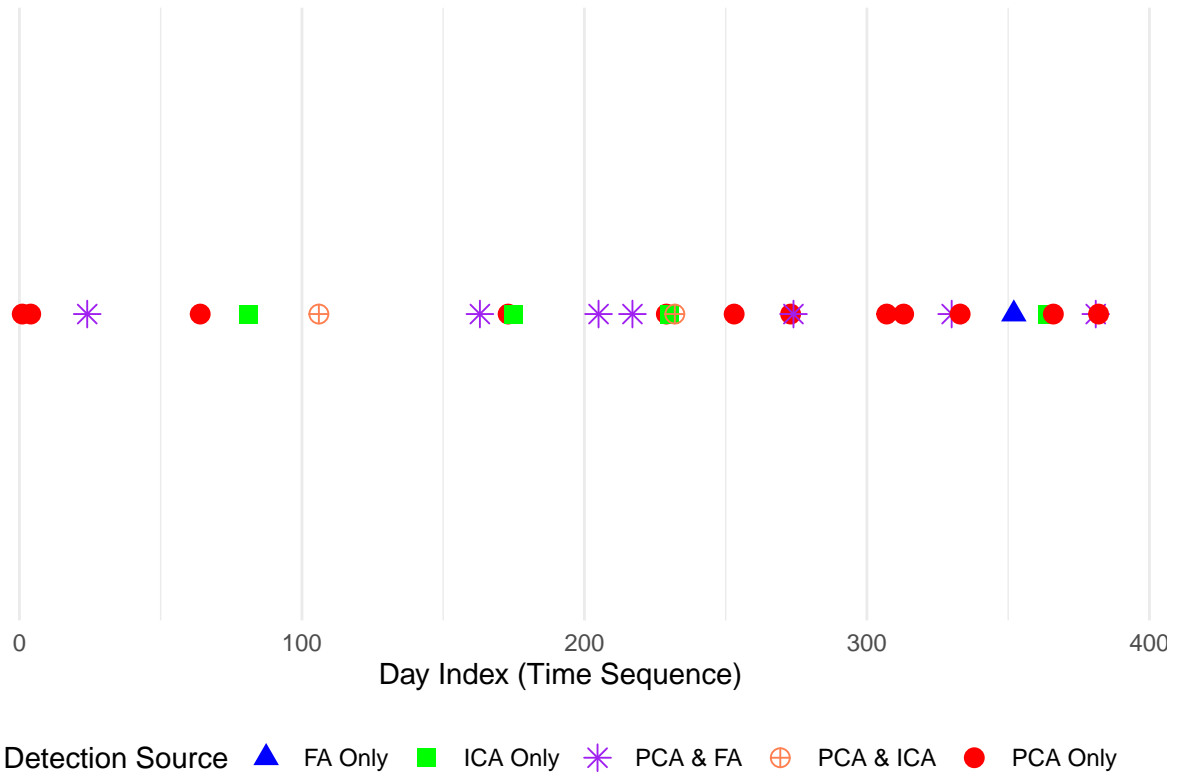
Anomaly Temporal Sequence – Loc12



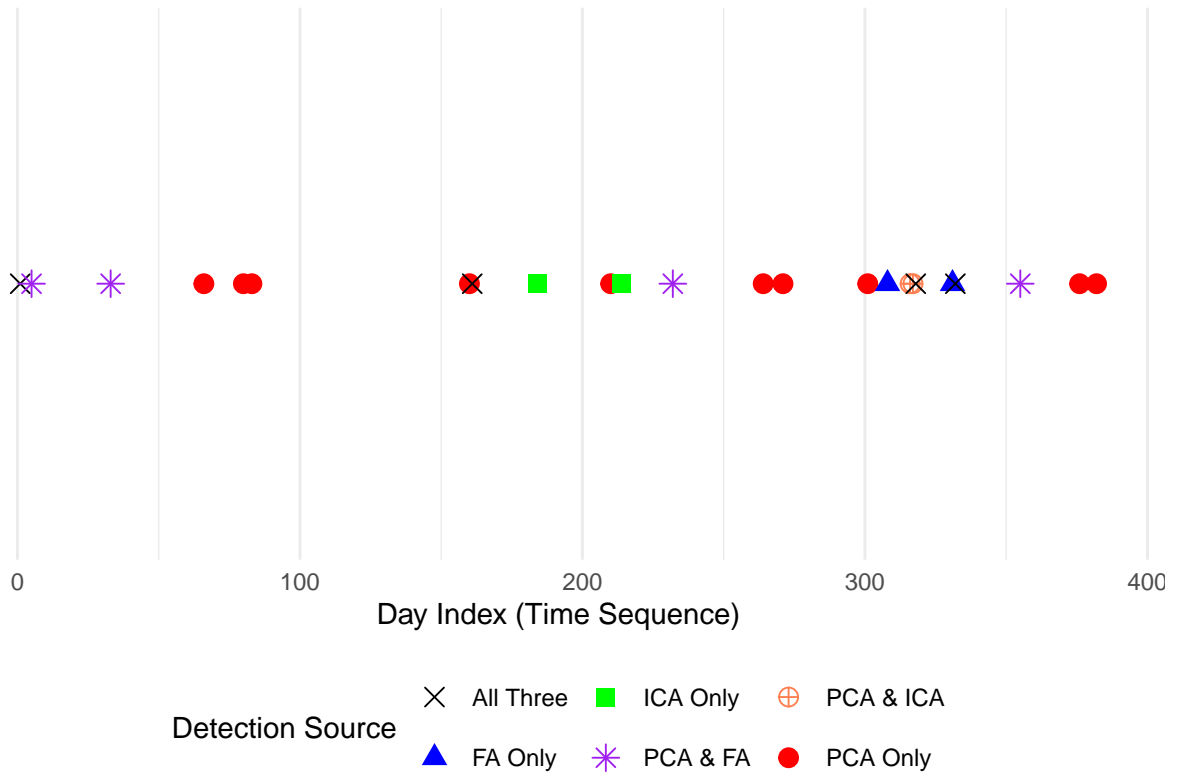
Anomaly Temporal Sequence – Loc13



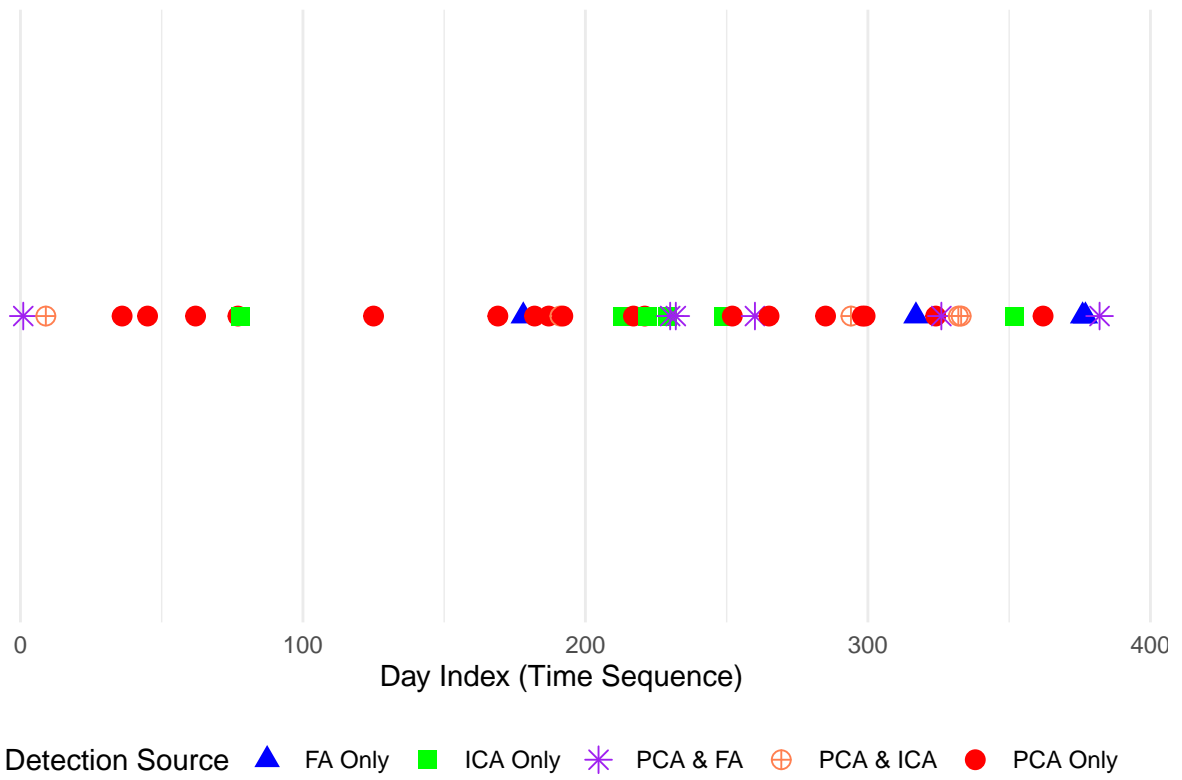
Anomaly Temporal Sequence – Loc14



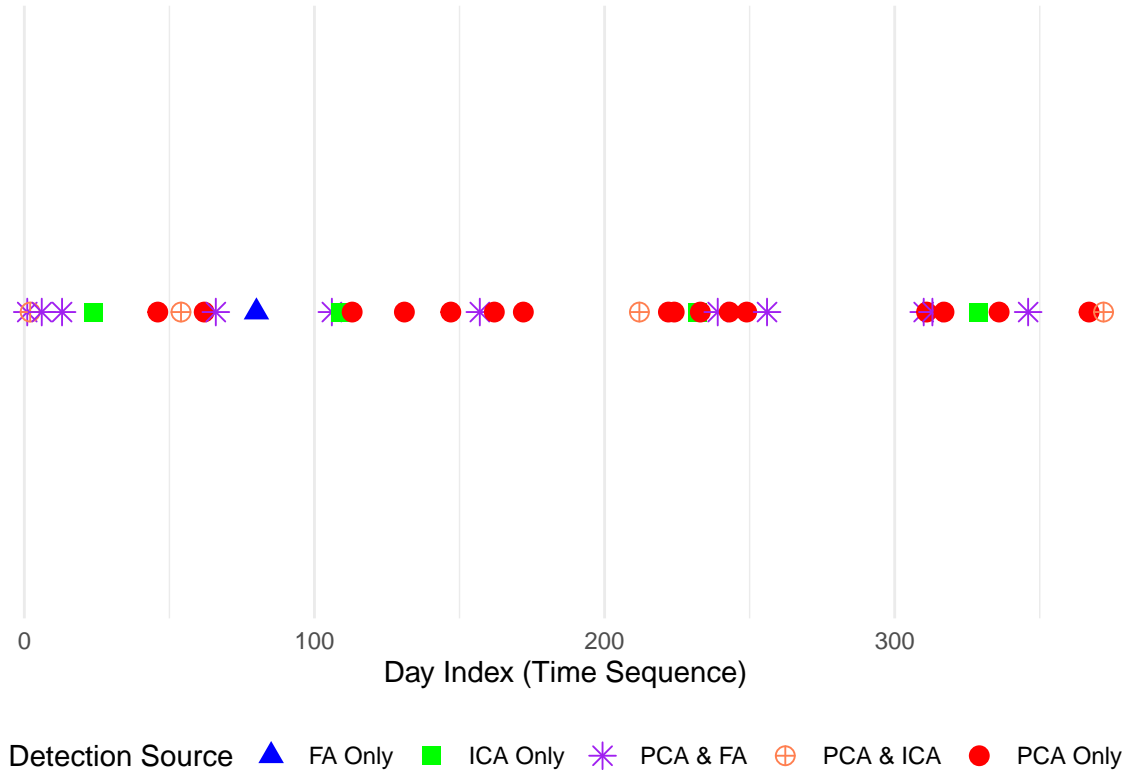
Anomaly Temporal Sequence – Loc15



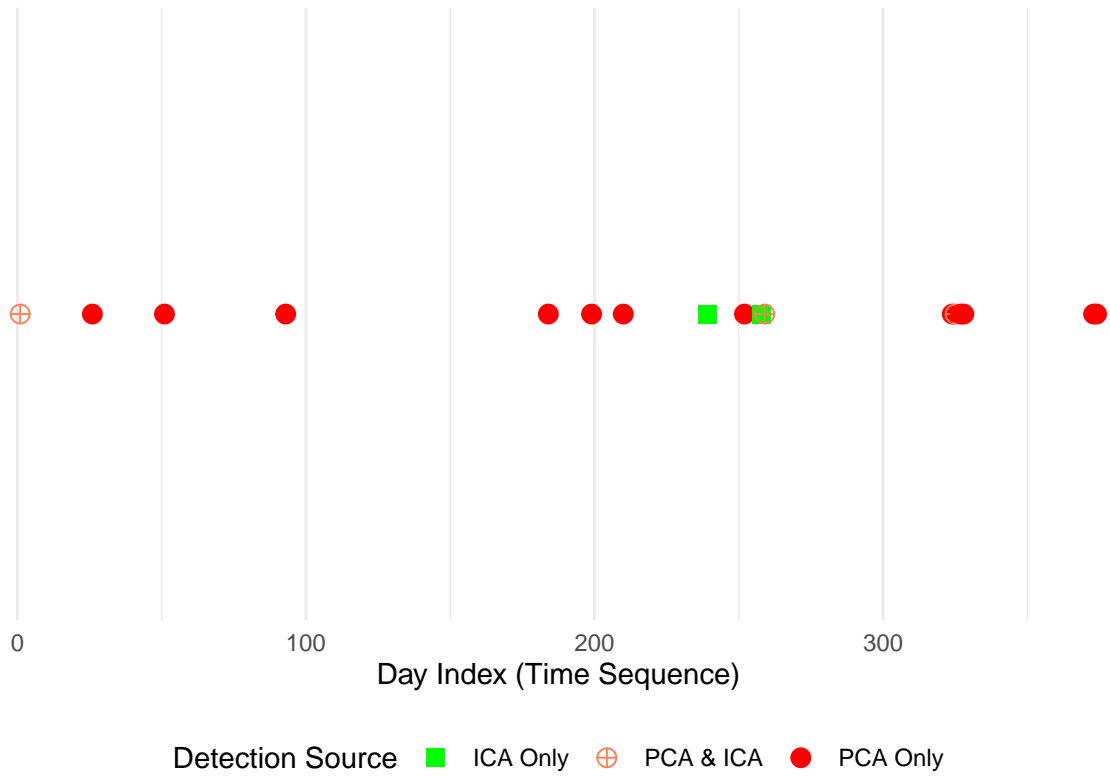
Anomaly Temporal Sequence – Loc16



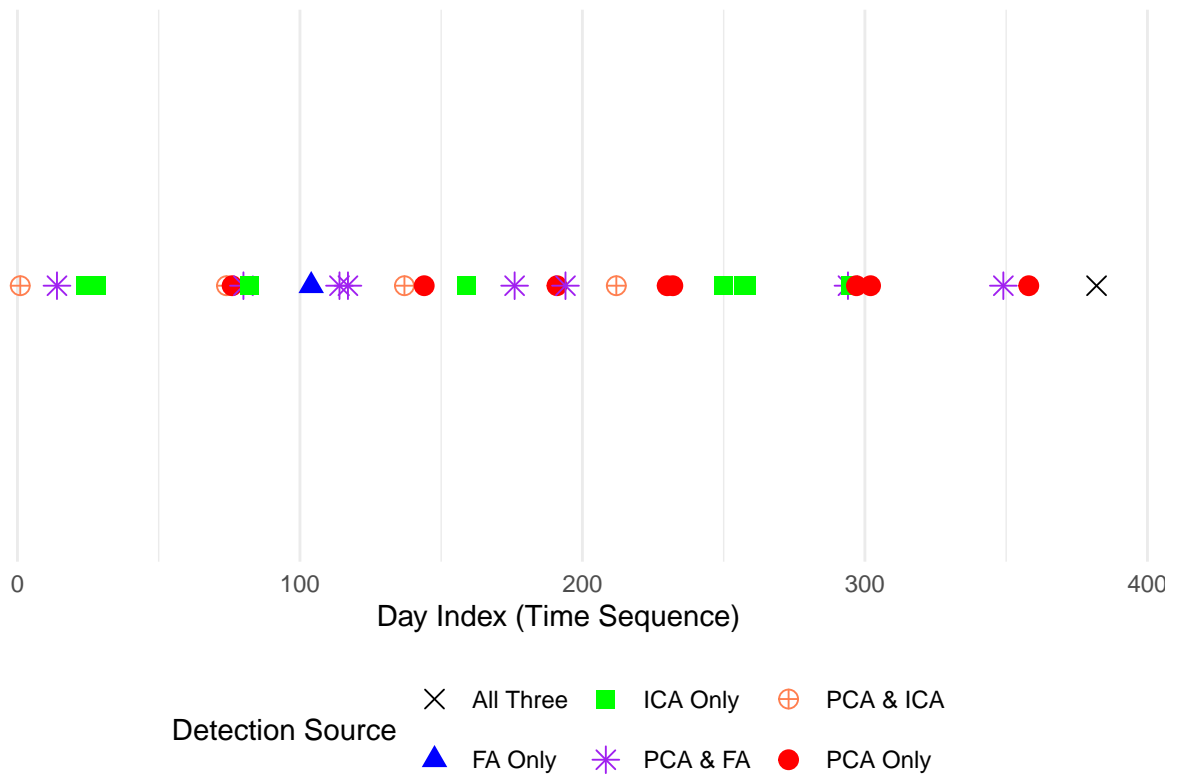
Anomaly Temporal Sequence – Loc17



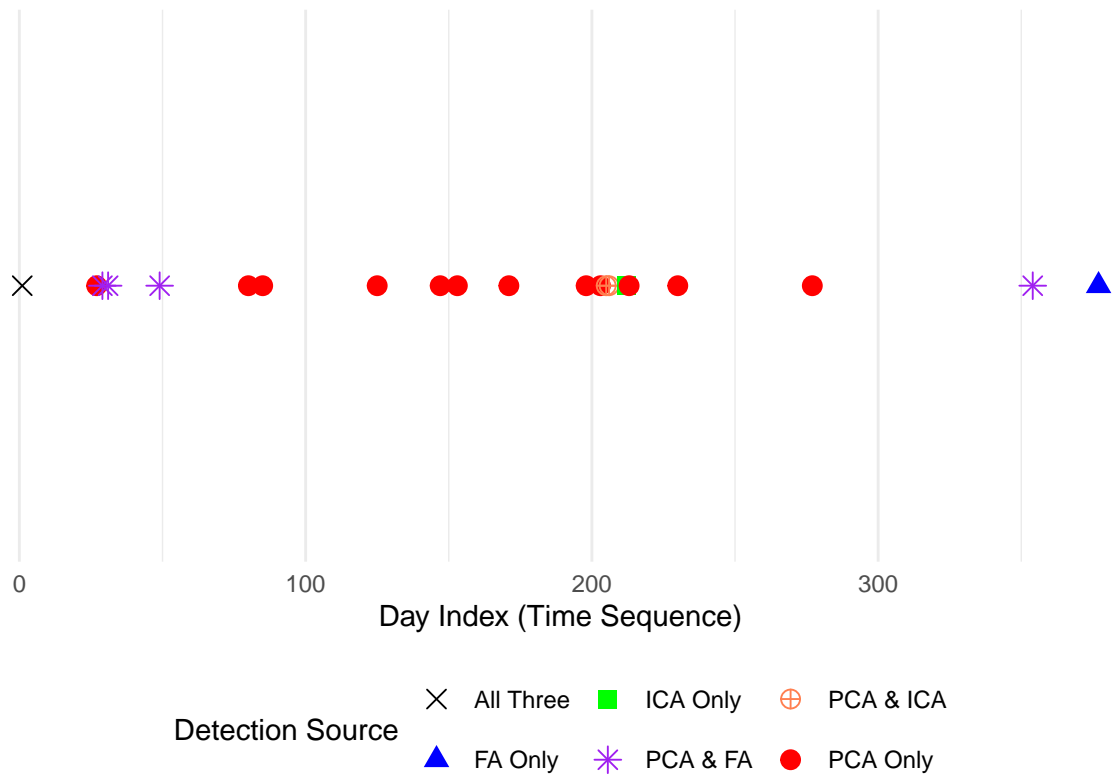
Anomaly Temporal Sequence – Loc18



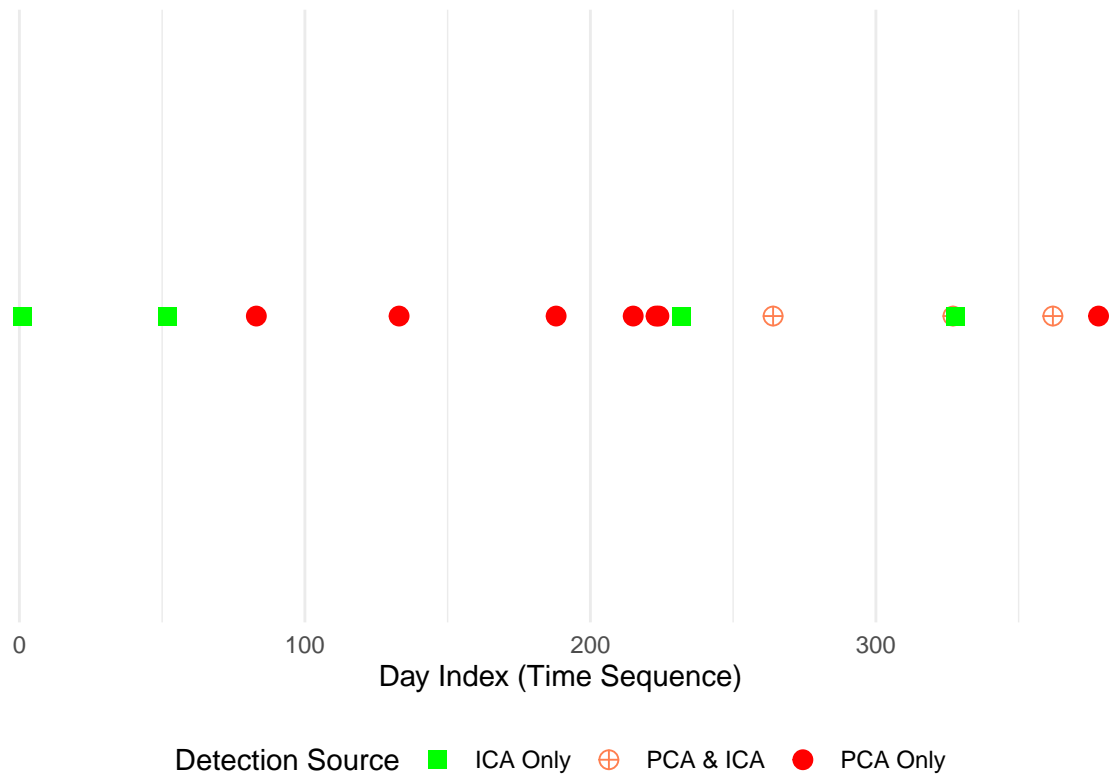
Anomaly Temporal Sequence – Loc19



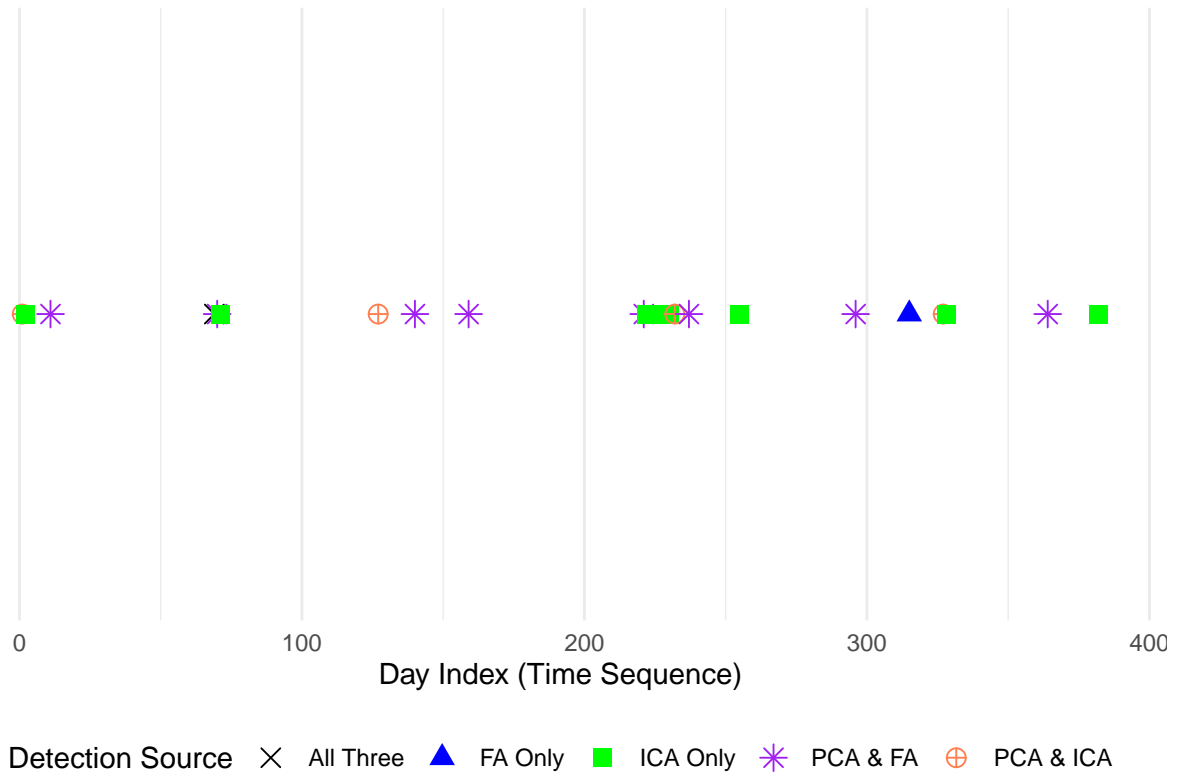
Anomaly Temporal Sequence – Loc20



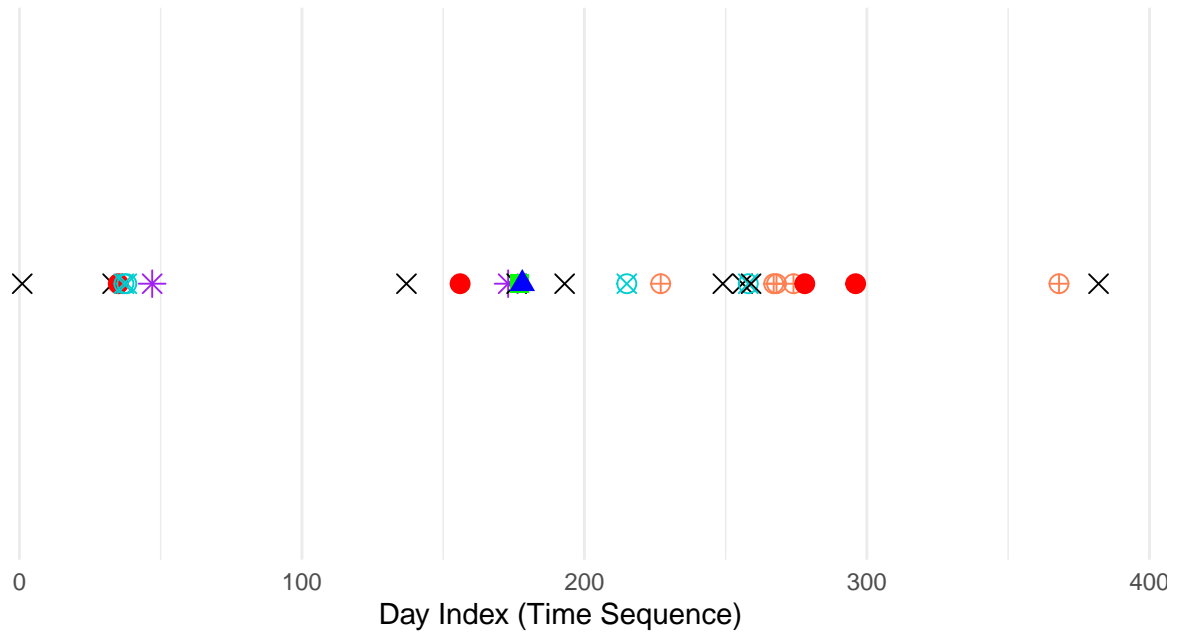
Anomaly Temporal Sequence – Loc21



Anomaly Temporal Sequence – Loc22



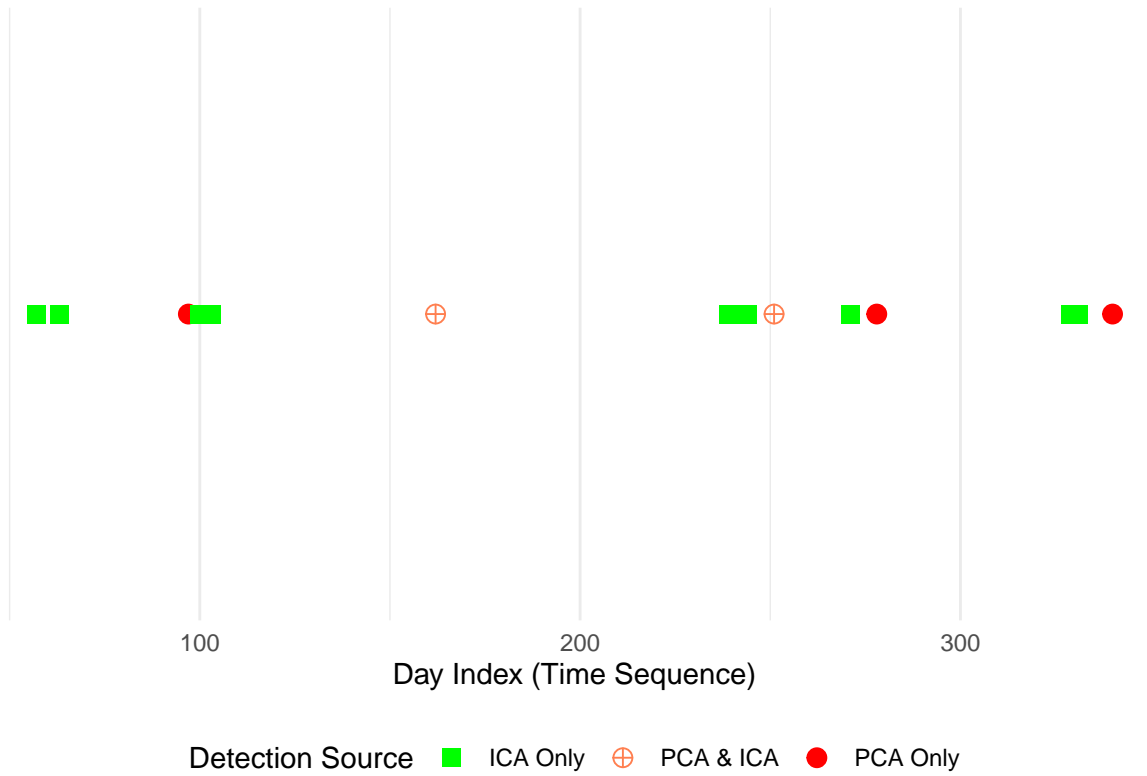
Anomaly Temporal Sequence – Loc23



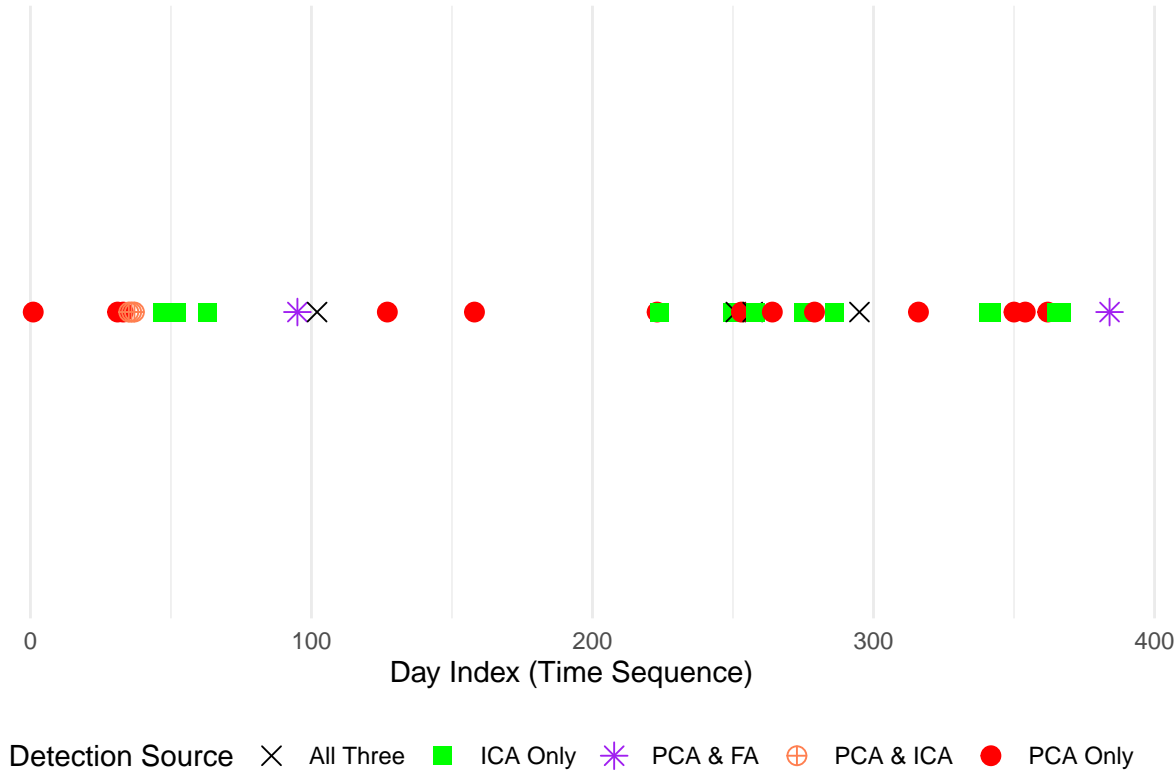
Detection Source

×	All Three	▲	FA Only	✱	PCA & FA	●	PCA Only
⊠	FA & ICA	■	ICA Only	⊕	PCA & ICA		

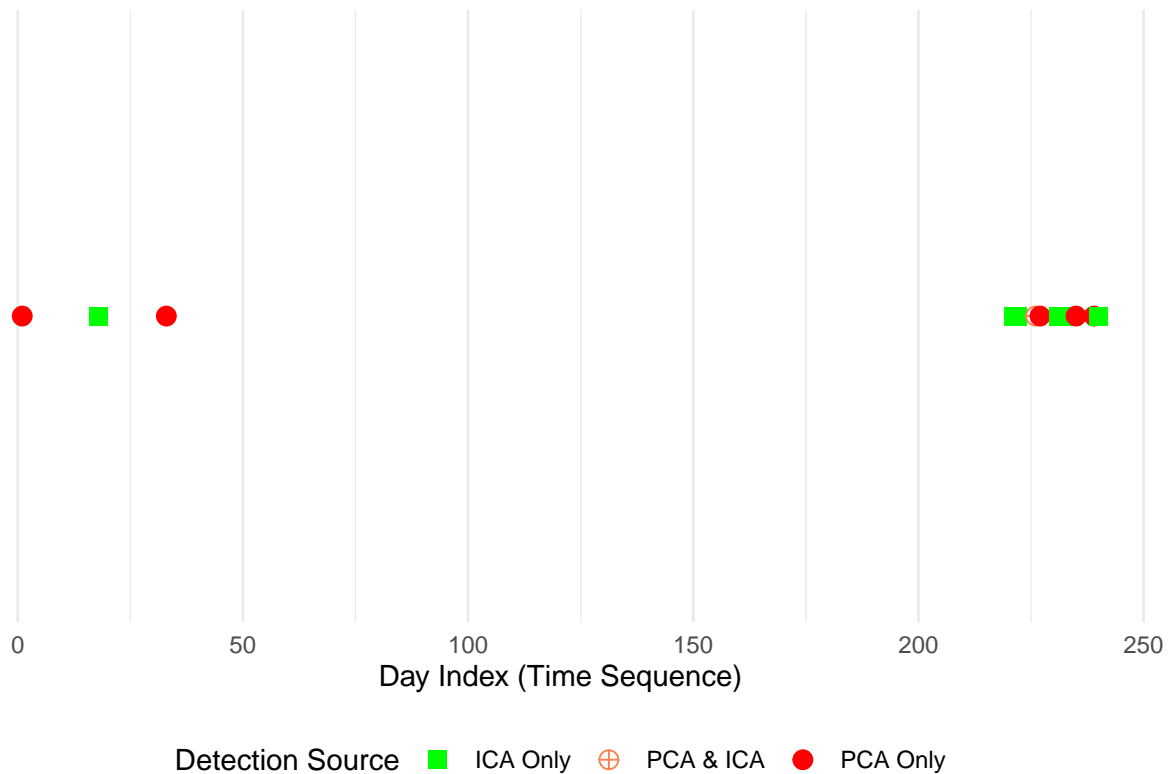
Anomaly Temporal Sequence – Loc24



Anomaly Temporal Sequence – Loc25



Anomaly Temporal Sequence – Loc26



ICA boxplots

```
## ICA Source Score Boxplot with Anomalies
plot_ica_score_boxplot <- function(res, loc_name, anomaly_coef = 1.5) {
  # Use res$ica_scores NOT res$ica$S!
  if (is.null(res$ica_scores) || nrow(res$ica_scores) == 0) {
    message("ICA scores not available or empty for ", loc_name)
    return(NULL)
  }

  scores <- res$ica_scores # ← CHANGED FROM res$ica$S
  days <- rownames(scores)

  # Verify rownames exist
  if (is.null(days)) {
    stop("ICA scores missing row names for ", loc_name)
  }

  # ICA anomalies (already calculated)
```

```

ica_anom_days <- res$ica_anom_days

# Debug
cat("Location:", loc_name, "- ICA Anomalies detected:", length(ica_anom_days), "\n")
if (length(ica_anom_days) > 0) {
  cat("  Anomaly days:", paste(head(ica_anom_days), collapse = ", "), "\n")
}

# Create the long data frame
df_long <- data.frame(
  Day = rep(days, ncol(scores)),
  Source = rep(paste0("Source", 1:ncol(scores)), each = nrow(scores)),
  Score = c(scores),
  stringsAsFactors = FALSE
)

# Identify and label anomalies
df_long$AnomalyType <- ifelse(df_long$Day %in% ica_anom_days, "ICA Anomaly", "Normal")

# Debug
cat("  Days marked as anomalies:", sum(df_long$AnomalyType == "ICA Anomaly"), "\n\n")

# Create the plot - SEPARATE LAYERS
ggplot(df_long, aes(x = Source, y = Score)) +
  geom_boxplot(outlier.shape = NA) +

  # Normal points first (light and small)
  geom_jitter(data = subset(df_long, AnomalyType == "Normal"),
    color = "gray70", alpha = 0.3, width = 0.15, size = 1) +

  # ICA anomalies on top (larger and bright)
  geom_point(data = subset(df_long, AnomalyType == "ICA Anomaly"),
    aes(color = "ICA Anomaly"), size = 1.0, alpha = 0.9) +

  scale_color_manual(values = c("ICA Anomaly" = "purple")) +
  labs(title = paste("ICA Source Score Boxplots with Anomalies -", loc_name),
    subtitle = paste("Total anomalies detected:", length(ica_anom_days)),
    x = "Independent Component Source",
    y = "Source Score",
    color = "Detection Method") +
  theme_minimal() +
  theme(legend.position = "bottom")
}

```



```

# Loop over locations to generate the plots
for (loc_name in names(location_results)) {
  tryCatch({
    print(plot_ica_score_boxplot(location_results[[loc_name]], loc_name))
  }, error = function(e) {
    warning(paste("Failed to plot ICA Score Boxplot for", loc_name, ":", e$message))
  })
}

```

FA loadings

```

## FA Factor Loadings Plot
plot_fa_loadings <- function(res, loc_name) {
  if (is.null(res$fa_loadings)) {
    message("FA results or loadings not available for ", loc_name)
    return(NULL)
  }

  loadings <- res$fa_loadings
  n_factors <- ncol(loadings)

  # Convert time intervals to hours (288 intervals = 24 hours)
  time_hours <- seq(0, 24, length.out = nrow(loadings))

  df <- data.frame(Time = time_hours, loadings)
  colnames(df) <- c("Time", paste0("Factor", 1:n_factors))

  df_long <- tidyr::pivot_longer(df, cols = -Time,
                                names_to = "Factor",
                                values_to = "Loading")

  ggplot(df_long, aes(x = Time, y = Loading, color = Factor)) +
    geom_line(size = 1) +
    facet_wrap(~Factor, ncol = 1, scales = "free_y") +
    labs(title = paste("FA Factor Loadings (Time Profile) -", loc_name),
         x = "Hour of Day",
         y = "Loading/Factor Weight") +
    theme_minimal() +
    scale_x_continuous(breaks = seq(0, 24, 4))
}

# Loop over locations to generate the plots

```

```
for (loc_name in names(location_results)) {
  print(plot_fa_loadings(location_results[[loc_name]], loc_name))
}
```

Venn diagrams we are not using this

```
plot_anomaly_venn<- function(res, loc_name) {
  anomaly_sets <- list(
    PCA = res$pca_anom_days,
    FA  = res$fa_anom_days,
    ICA = res$ica_anom_days
  )

  # Remove empty sets
  anomaly_sets <- anomaly_sets[sapply(anomaly_sets, length) > 0]

  if(length(anomaly_sets) < 2) {
    message("Not enough methods for Venn at ", loc_name)
    return(NULL)
  }

  ggvenn(
    anomaly_sets,
    fill_color = c("#FF6B6B", "#4D96FF", "#9B5DE5"),
    fill_alpha = 0.55,
    stroke_size = 0.8,
    text_size = 6,           # Much larger text
    show_percentage = FALSE # Turn off % labels
  ) +
  labs(title = paste("Anomaly Overlap -", loc_name)) +
  theme(plot.title = element_text(hjust = 0.5, size = 14, face = "bold"))
}

# Loop over locations
for (loc_name in names(location_results)) {
  tryCatch({
    print(plot_anomaly_venn(location_results[[loc_name]], loc_name))
  }, error = function(e) {
    warning(paste("Failed to plot Venn diagram for", loc_name, ":", e$message))
  })
}
```

Cluster

```
# Step 1: Get the minimum number of retained PCs across all locations
min_k <- min(sapply(location_results, function(res) res$k_pca))
cat("Minimum number of available PCs across locations:", min_k, "\n")
```

```
## Minimum number of available PCs across locations: 2
```

```
# Step 2: Extract ONLY the first min_k PCs from each location
pca_all_scores <- do.call(rbind, lapply(location_results, function(res) {
  if (ncol(res$pca_scores) >= min_k) {
    res$pca_scores[, 1:min_k, drop = FALSE]
  } else {
    NULL # In case some failed completely
  }
}))

# Optional: remove duplicate row names
pca_all_scores <- pca_all_scores[!duplicated(rownames(pca_all_scores)), ]

# Step 3: Compute silhouette scores for different K

silhouette_scores <- data.frame(K = integer(), Silhouette = double())

for (k in 2:6) {
  km <- kmeans(scale(pca_all_scores), centers = k, nstart = 25)
  sil <- silhouette(km$cluster, dist(scale(pca_all_scores)))
  silhouette_scores <- rbind(silhouette_scores,
                             data.frame(K = k, Silhouette = mean(sil[, 3])))
}

print(silhouette_scores)
```

```
##    K Silhouette
## 1 2  0.3098061
## 2 3  0.3424861
## 3 4  0.3226218
## 4 5  0.3302555
## 5 6  0.3611448
```

```
best_k <- silhouette_scores$K[which.max(silhouette_scores$Silhouette)]
cat("Best number of clusters based on silhouette:", best_k, "\n")
```

```
## Best number of clusters based on silhouette: 6
```

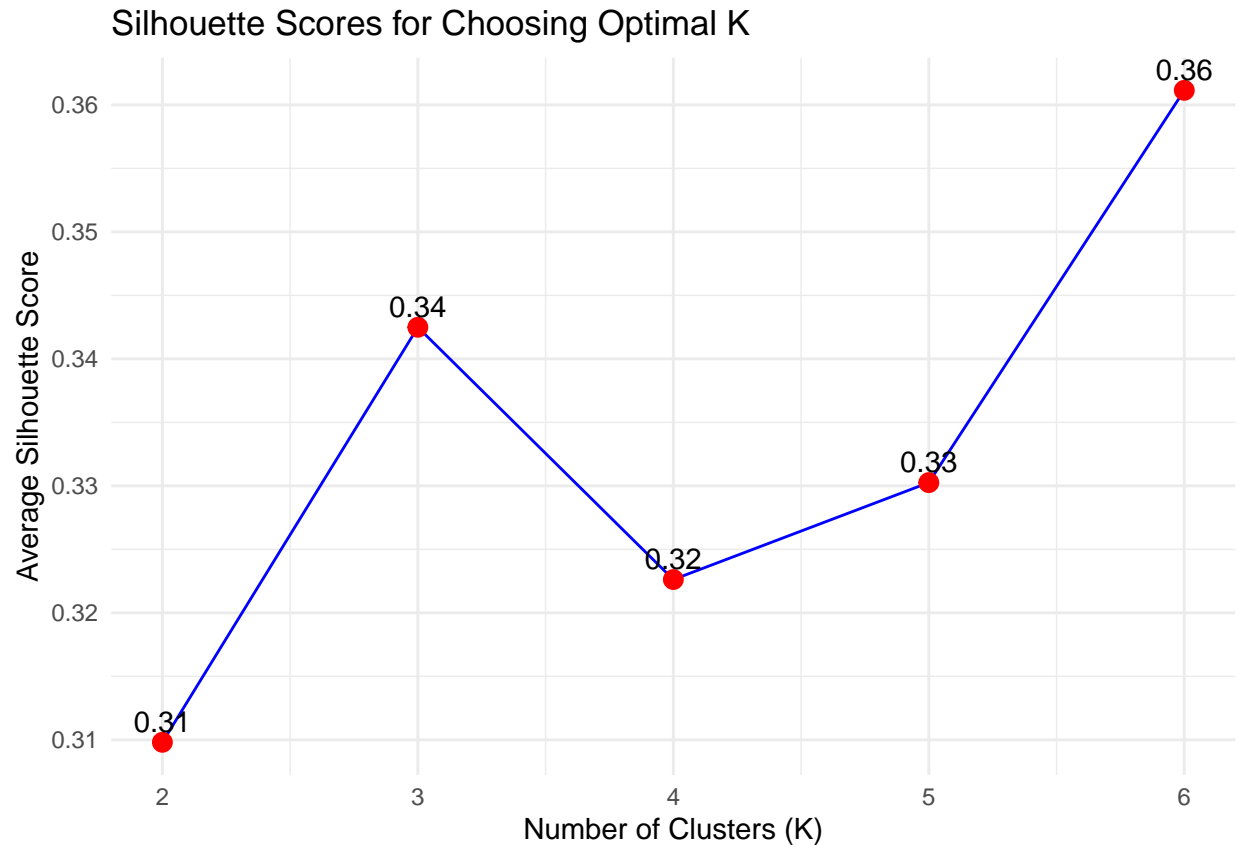
```
km_res <- kmeans(scale(pca_all_scores), centers = best_k, nstart = 25)

cluster_labels <- data.frame(
  Day = rownames(pca_all_scores),
  Cluster = as.factor(km_res$cluster)
)

head(cluster_labels)
```

```
##           Day Cluster
## WkDay-1 WkDay-1      5
## WkDay-2 WkDay-2      5
## WkDay-3 WkDay-3      5
## WkDay-4 WkDay-4      6
## WkDay-5 WkDay-5      3
## WkDay-6 WkDay-6      2
```

```
ggplot(silhouette_scores, aes(x = K, y = Silhouette)) +
  geom_line(color = "blue") +
  geom_point(size = 3, color = "red") +
  geom_text(aes(label = round(Silhouette, 2)), vjust = -0.5) +
  labs(title = "Silhouette Scores for Choosing Optimal K",
       x = "Number of Clusters (K)",
       y = "Average Silhouette Score") +
  theme_minimal()
```

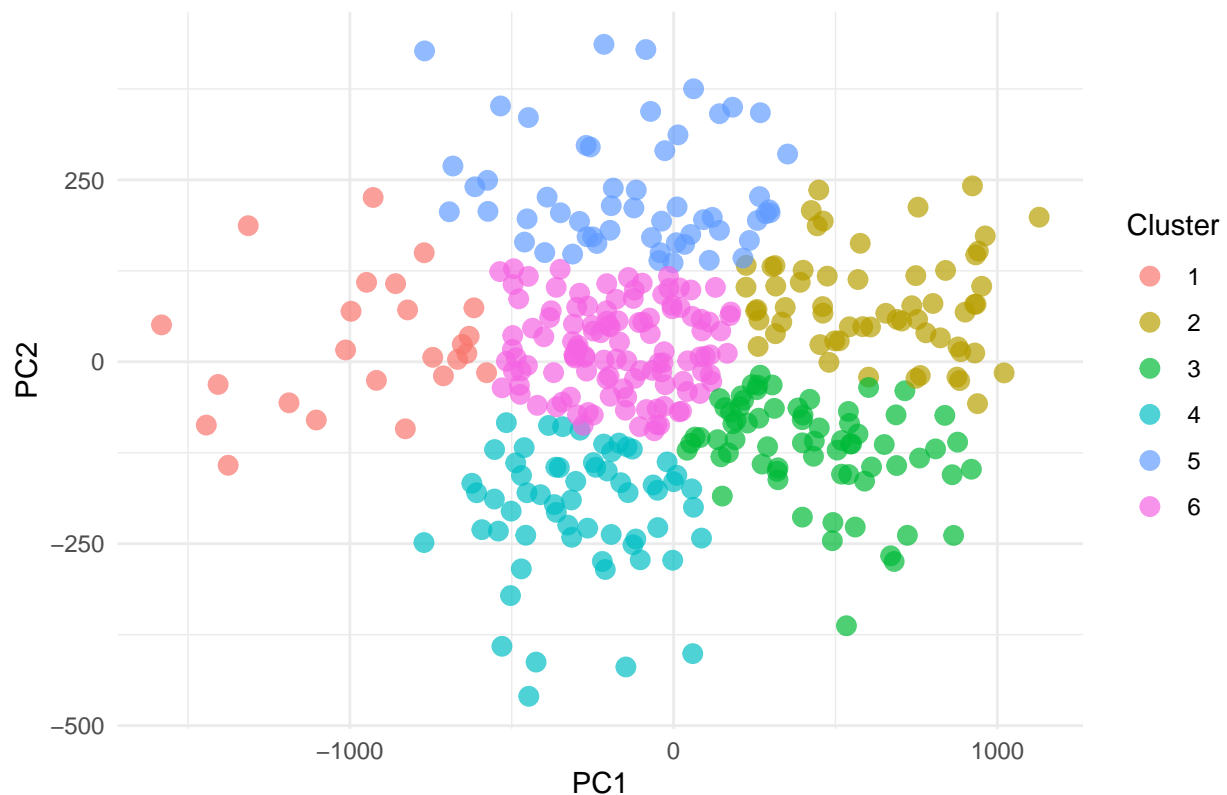


Plot PCA Score Scatterplot Colored by Cluster

This shows how clusters separate in feature space, not along time.

```
pca_cluster_plot <- function(pca_all_scores, cluster_labels) {  
  df <- data.frame(pca_all_scores[, 1:2], Cluster = cluster_labels$Cluster)  
  colnames(df)[1:2] <- c("PC1", "PC2")  
  
  ggplot(df, aes(x = PC1, y = PC2, color = Cluster)) +  
    geom_point(size = 3, alpha = 0.7) +  
    labs(title = "Clusters Visualized in PCA Space",  
         x = "PC1",  
         y = "PC2",  
         color = "Cluster") +  
    theme_minimal()  
}  
  
pca_cluster_plot(pca_all_scores, cluster_labels)
```

Clusters Visualized in PCA Space

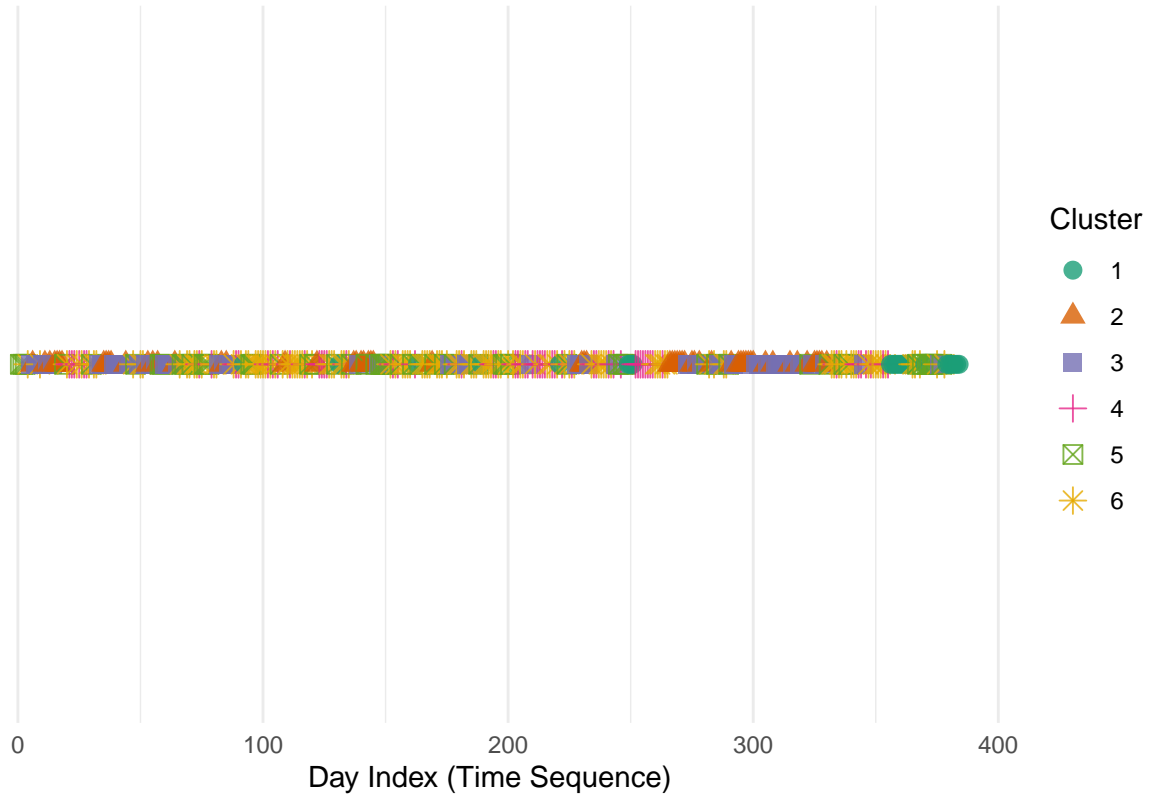


```
plot_cluster_timeline <- function(cluster_labels) {
  cluster_labels$DayIndex <- 1:nrow(cluster_labels) # Order by time

  ggplot(cluster_labels, aes(x = DayIndex, y = 1, color = Cluster, shape = Cluster)) +
    geom_point(size = 3, alpha = 0.8) +
    scale_color_brewer(palette = "Dark2") +
    labs(
      title = "Cluster Occurrences in Time (Sequential Days)",
      x = "Day Index (Time Sequence)",
      y = "",
      color = "Cluster",
      shape = "Cluster"
    ) +
    theme_minimal() +
    theme(
      axis.text.y = element_blank(),
      axis.ticks.y = element_blank(),
      panel.grid.major.y = element_blank(),
      panel.grid.minor.y = element_blank()
    )
}
```

```
# Example call:
plot_cluster_timeline(cluster_labels)
```

Cluster Occurrences in Time (Sequential Days)



```
plot_cluster_mean_shapes <- function(res, X, cluster_df, loc_name) {
  X_t <- t(X)

  df_long <- data.frame(
    Day = rownames(X_t),
    X_t,
    Cluster = cluster_df$Cluster[match(rownames(X_t), cluster_df$Day)]
  ) %>%
  filter(!is.na(Cluster)) %>%
  pivot_longer(cols = -c(Day, Cluster),
               names_to = "TimePoint",
               values_to = "Traffic") %>%
  mutate(TimeHour = (as.numeric(gsub("X", "", TimePoint)) - 1) * (24 / 287))

  df_mean <- df_long %>%
  group_by(Cluster, TimeHour) %>%
  summarise(MeanTraffic = mean(Traffic, na.rm = TRUE), .groups = "drop")
}
```

```

ggplot(df_mean, aes(x = TimeHour, y = MeanTraffic, color = Cluster)) +
  geom_line(size = 1.1) +
  labs(
    title = paste("Mean Traffic Curves by Cluster -", loc_name),
    x = "Hour of Day",
    y = "Mean Traffic Volume",
    color = "Cluster"
  ) +
  scale_x_continuous(breaks = seq(0, 24, 4)) +
  theme_minimal() +
  theme(legend.position = "bottom") +
  guides(color = guide_legend(position = "bottom"))
}

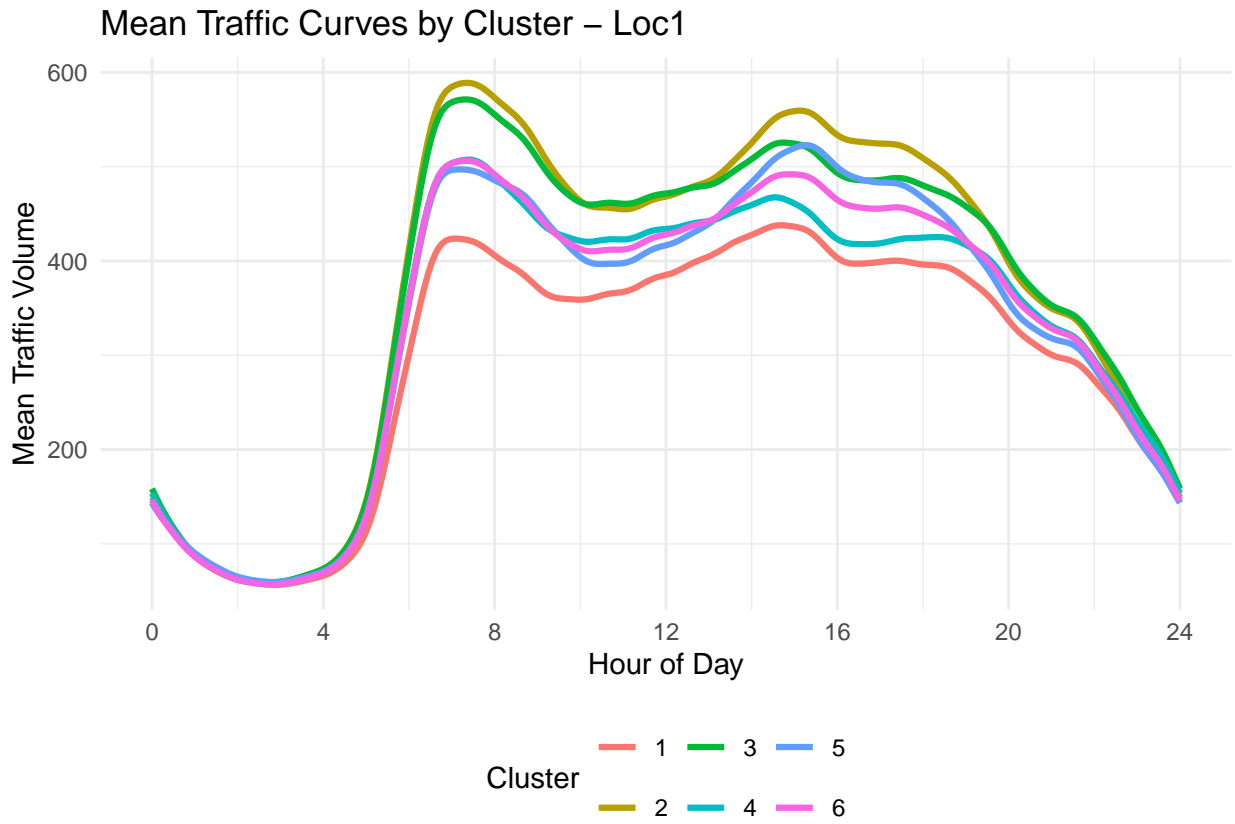
for (loc_name in names(location_results)) {
  print(
    plot_cluster_mean_shapes(
      res = location_results[[loc_name]], # model results for this location
      X = df[[loc_name]],                # original traffic data
      cluster_df = cluster_labels,        # full anomaly cluster labels
      loc_name = loc_name                  # title
    )
  )
}

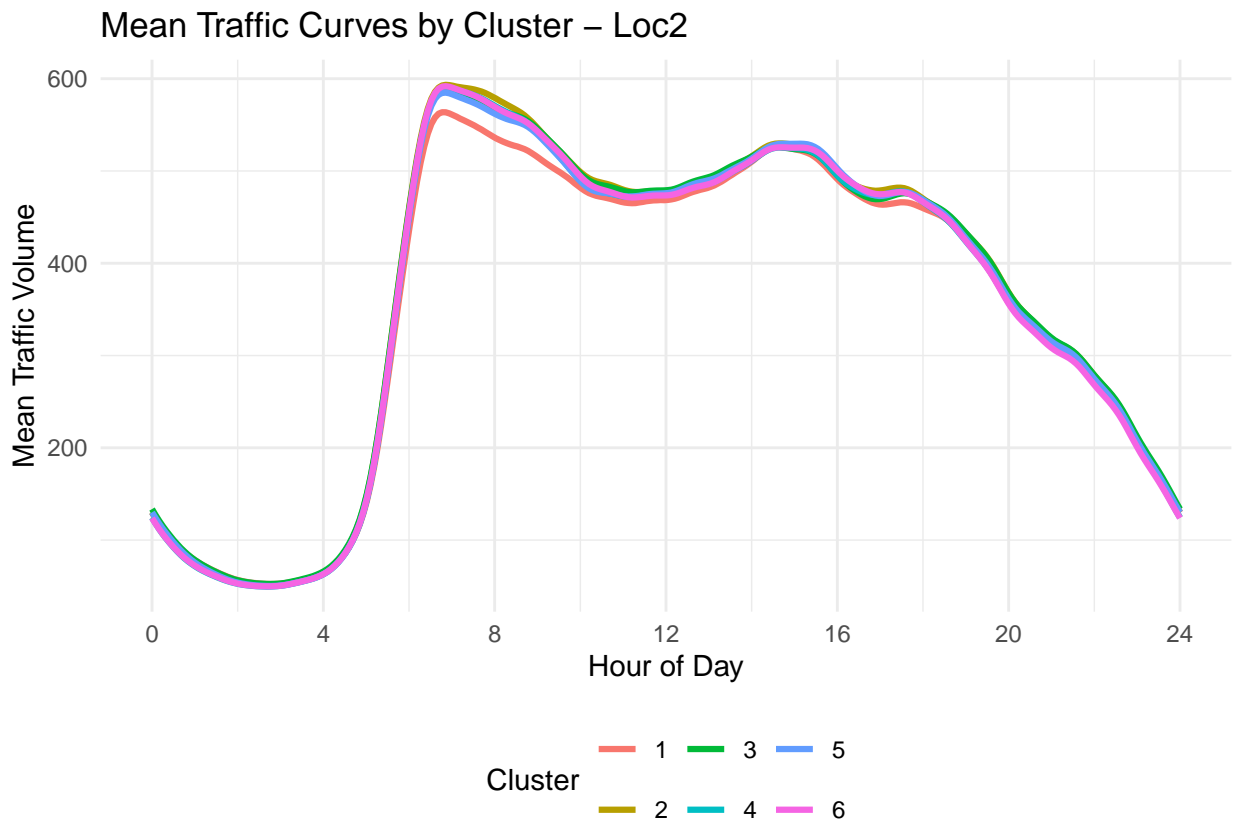
```

```

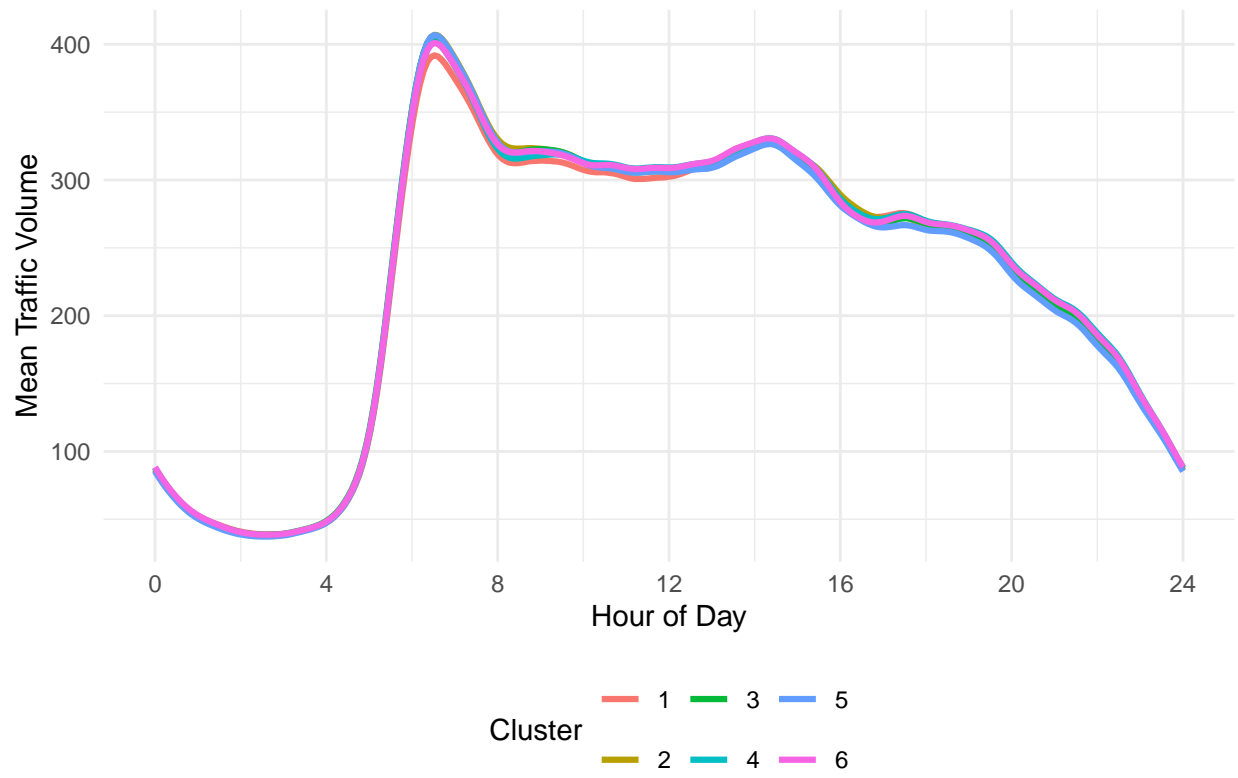
## Warning: Using 'size' aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use 'linewidth' instead.
## This warning is displayed once every 8 hours.
## Call 'lifecycle::last_lifecycle_warnings()' to see where this warning was
## generated.

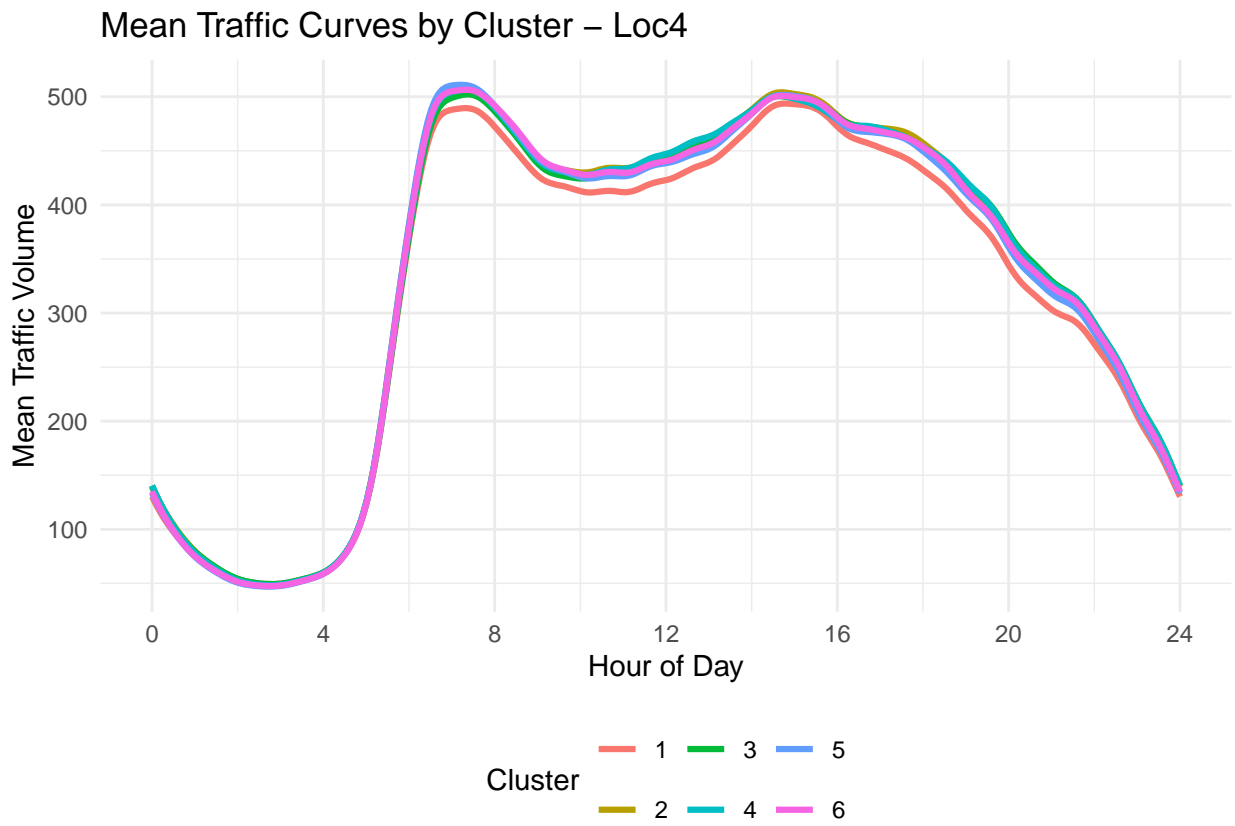
```

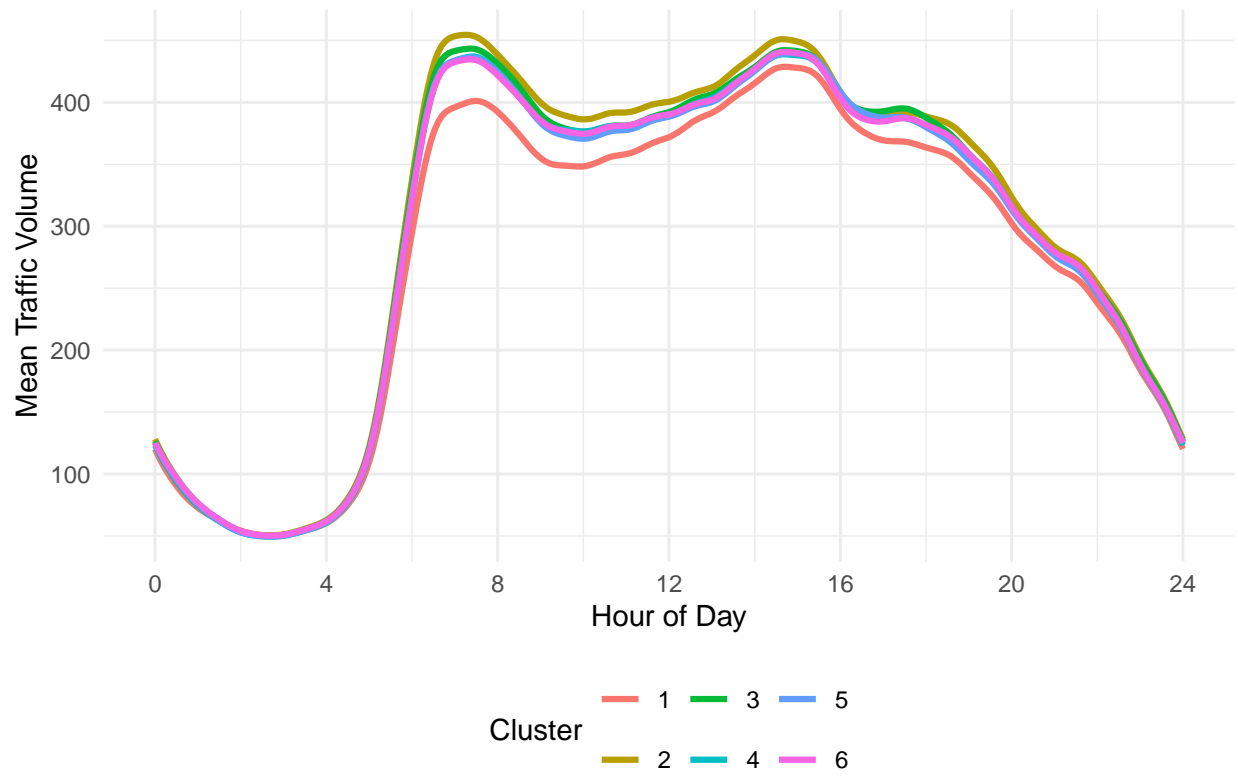


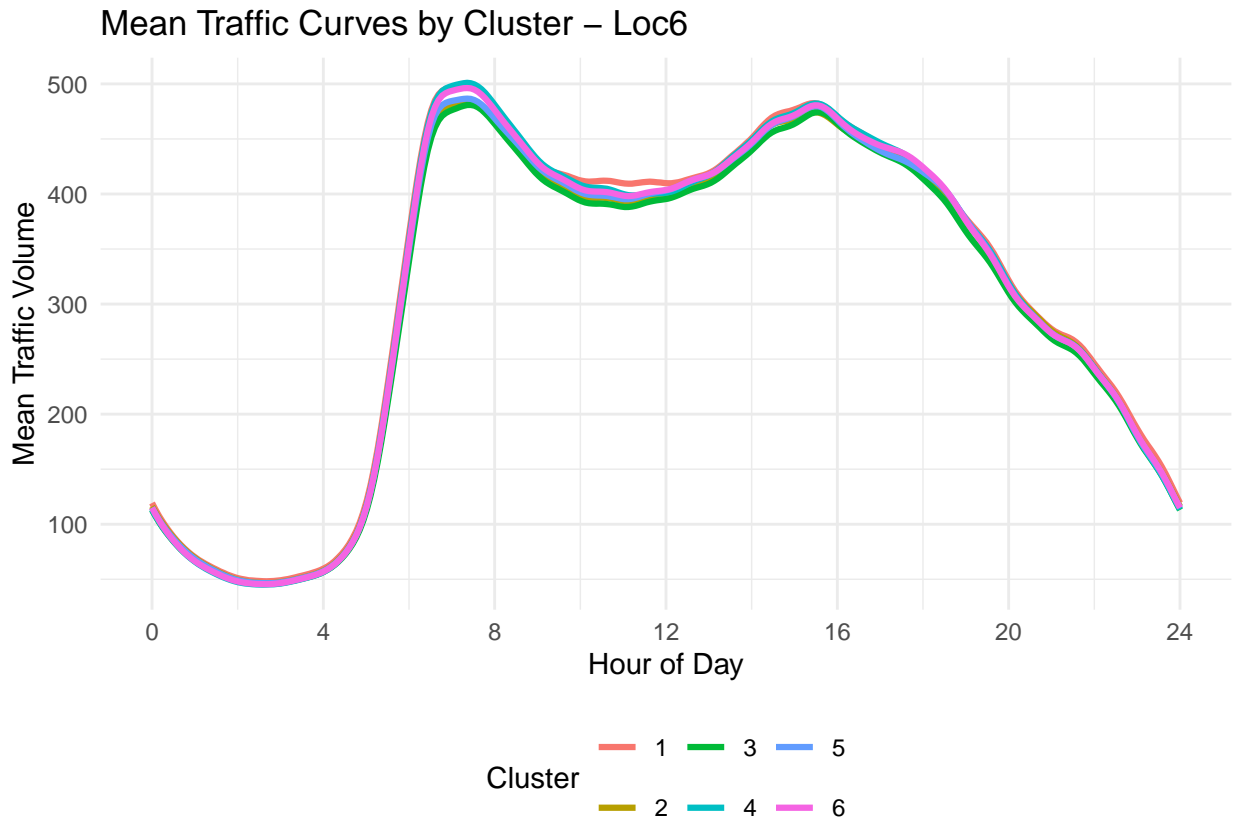
Mean Traffic Curves by Cluster – Loc3



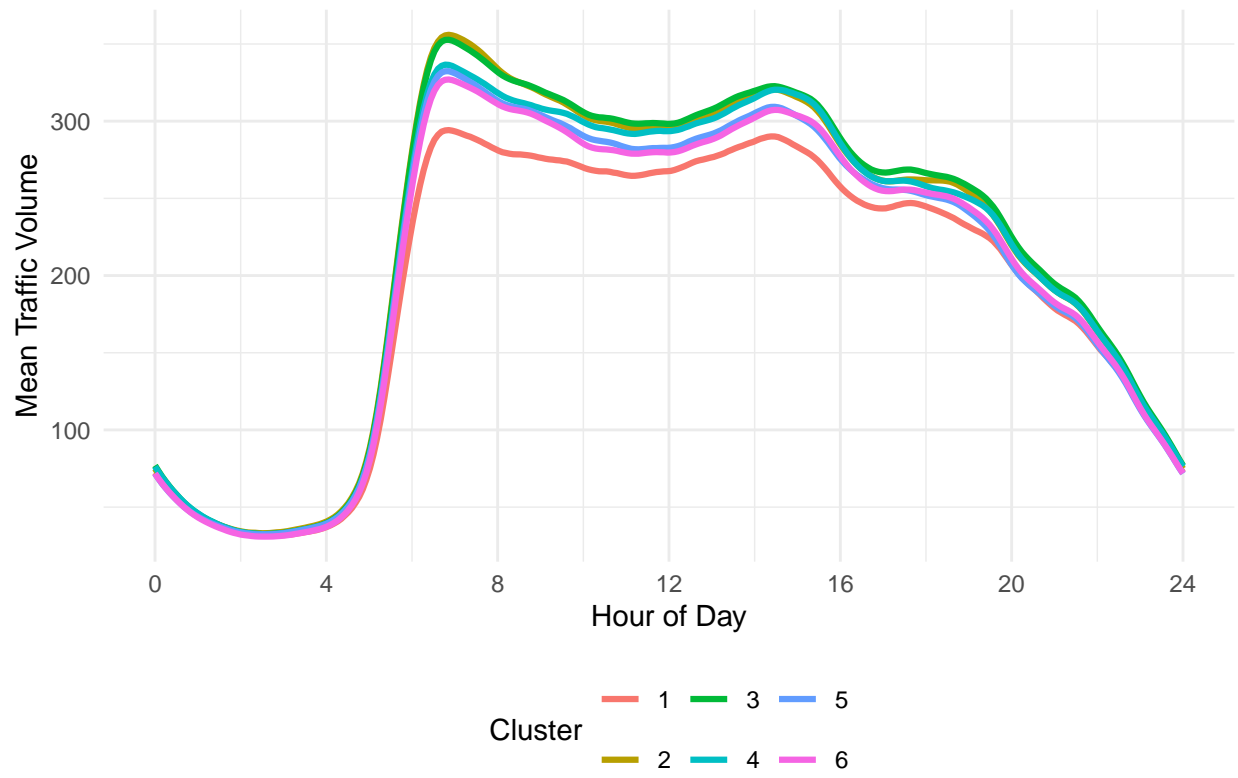


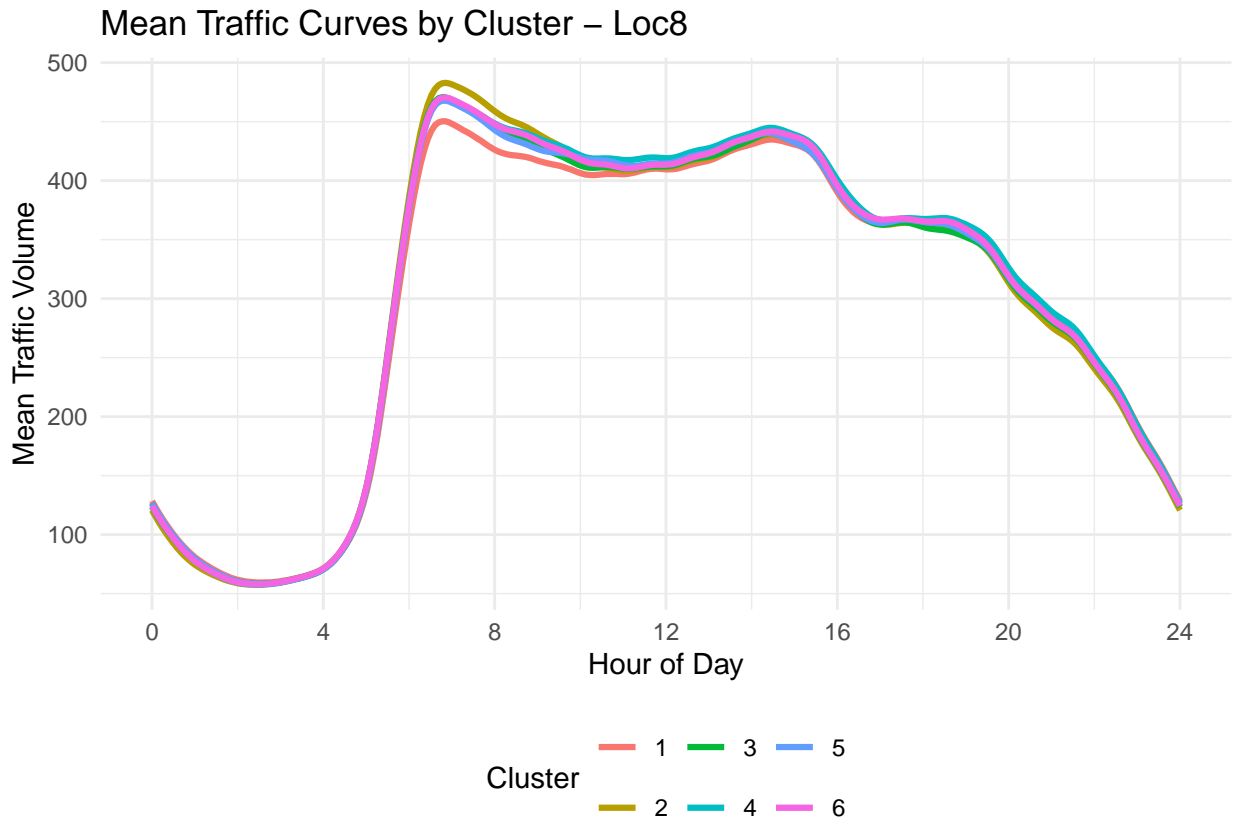
Mean Traffic Curves by Cluster – Loc5



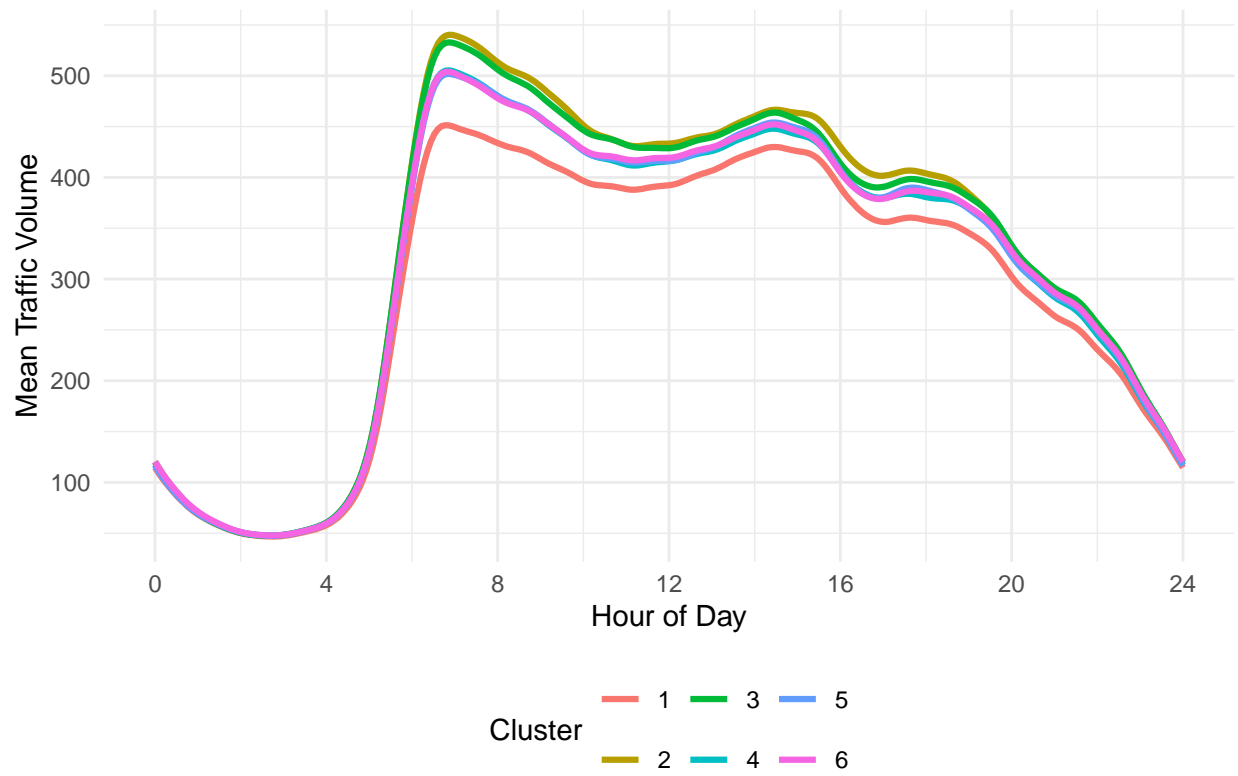


Mean Traffic Curves by Cluster – Loc7

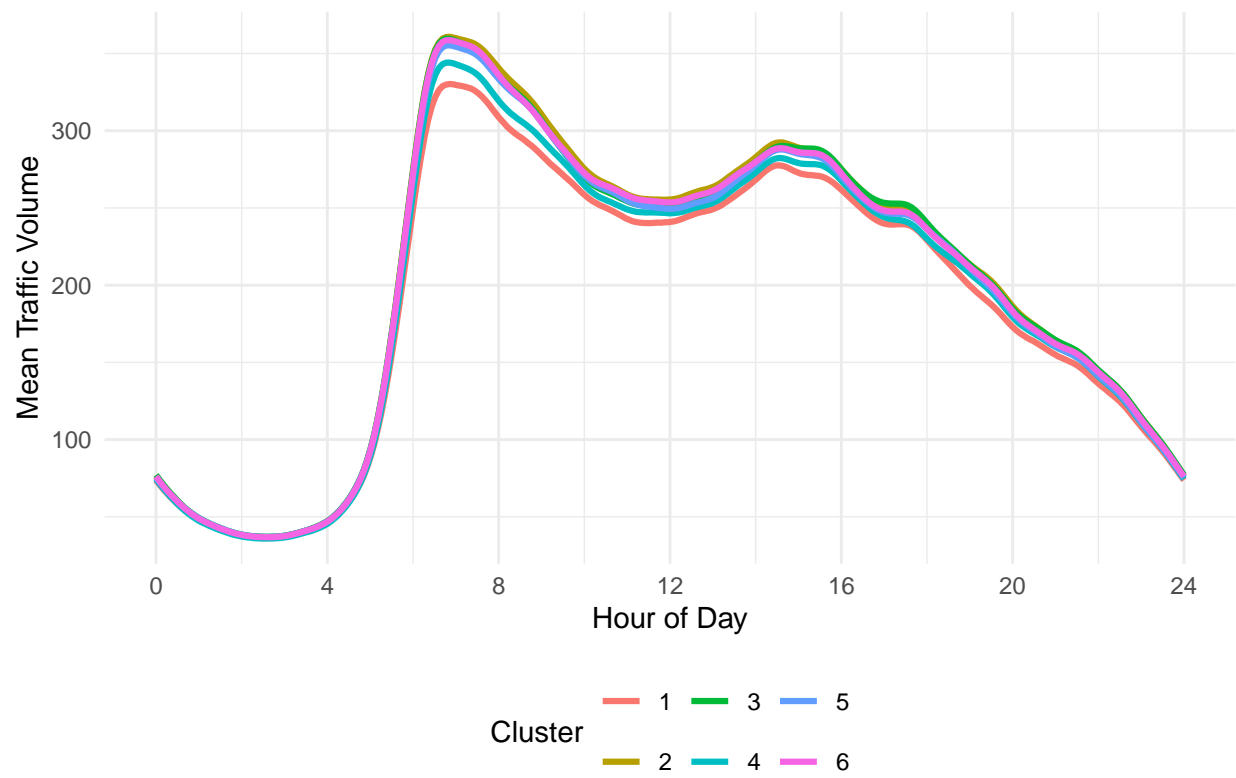


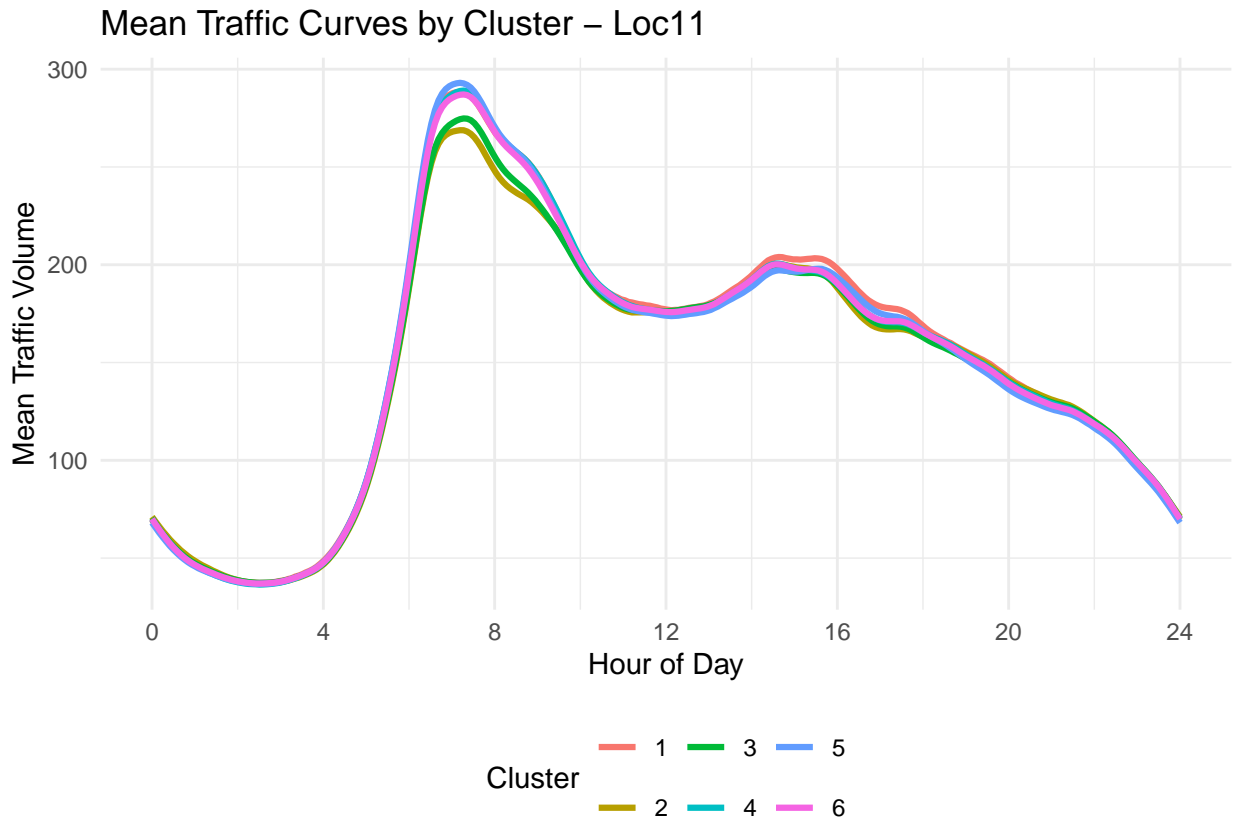


Mean Traffic Curves by Cluster – Loc9

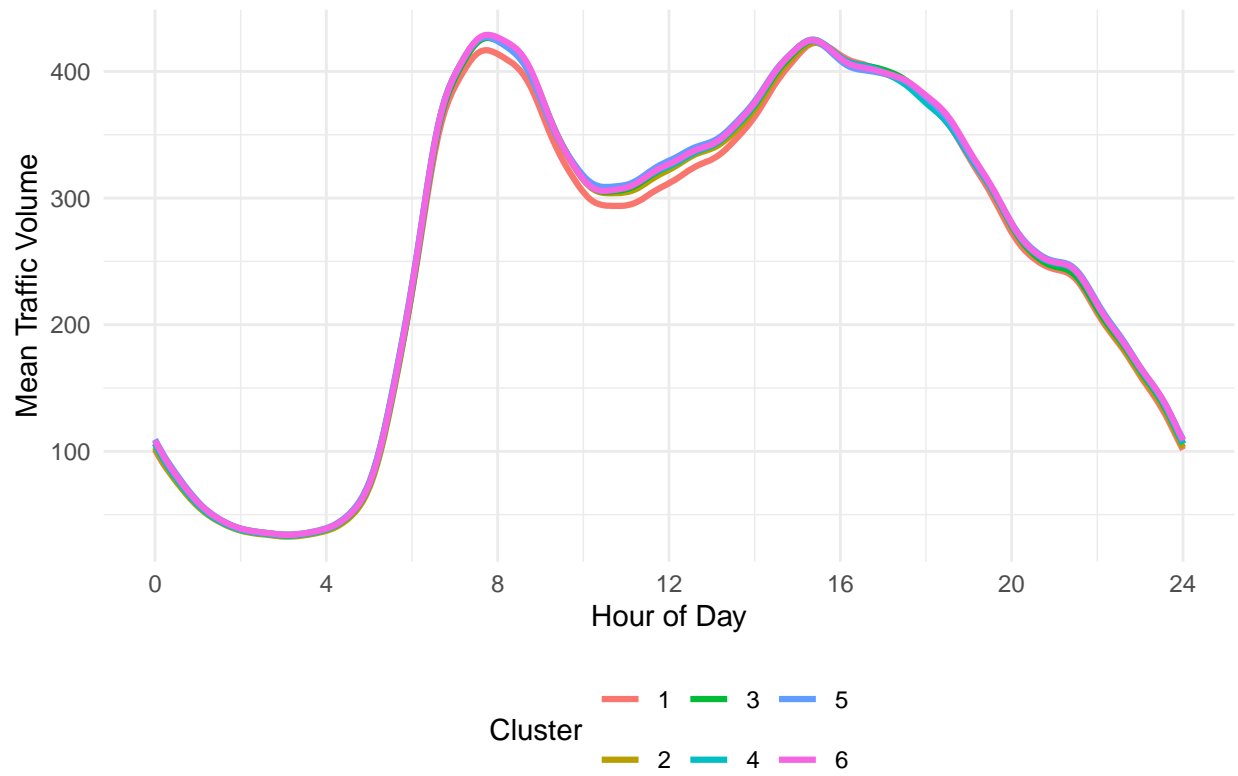


Mean Traffic Curves by Cluster – Loc10

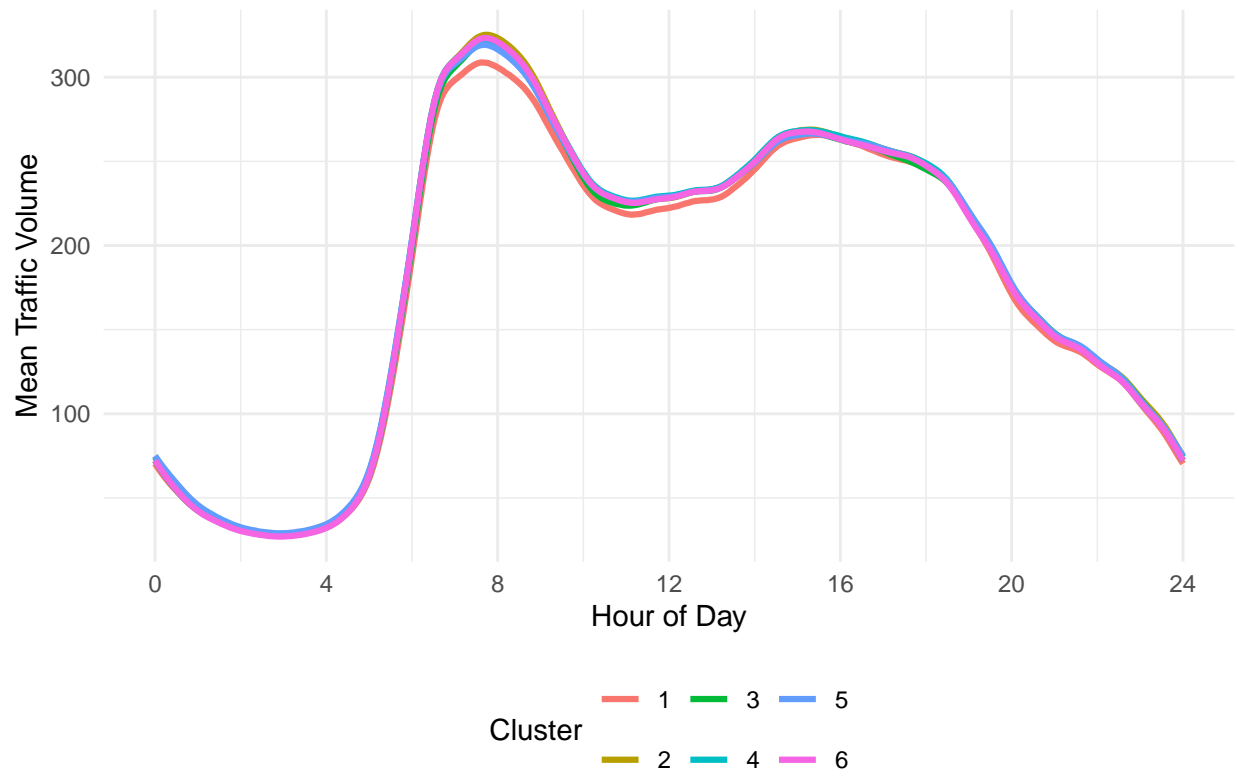


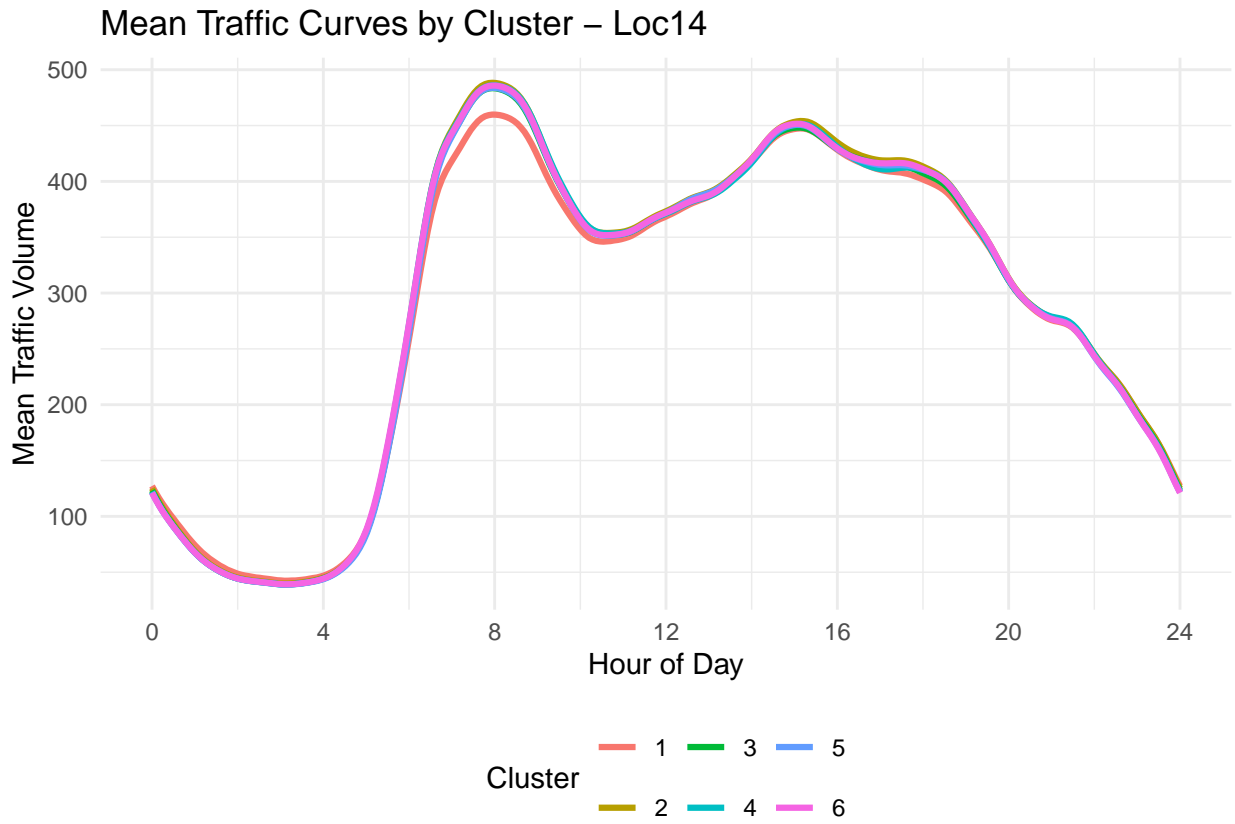


Mean Traffic Curves by Cluster – Loc12

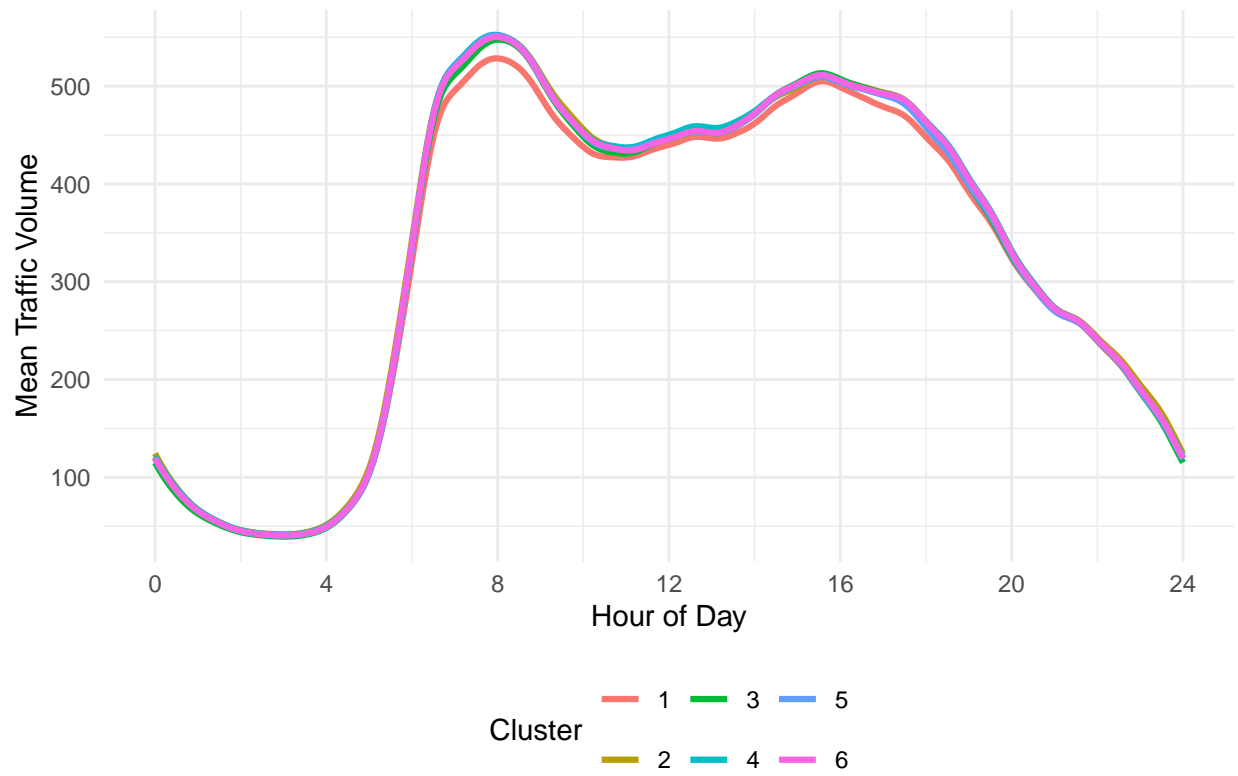


Mean Traffic Curves by Cluster – Loc13

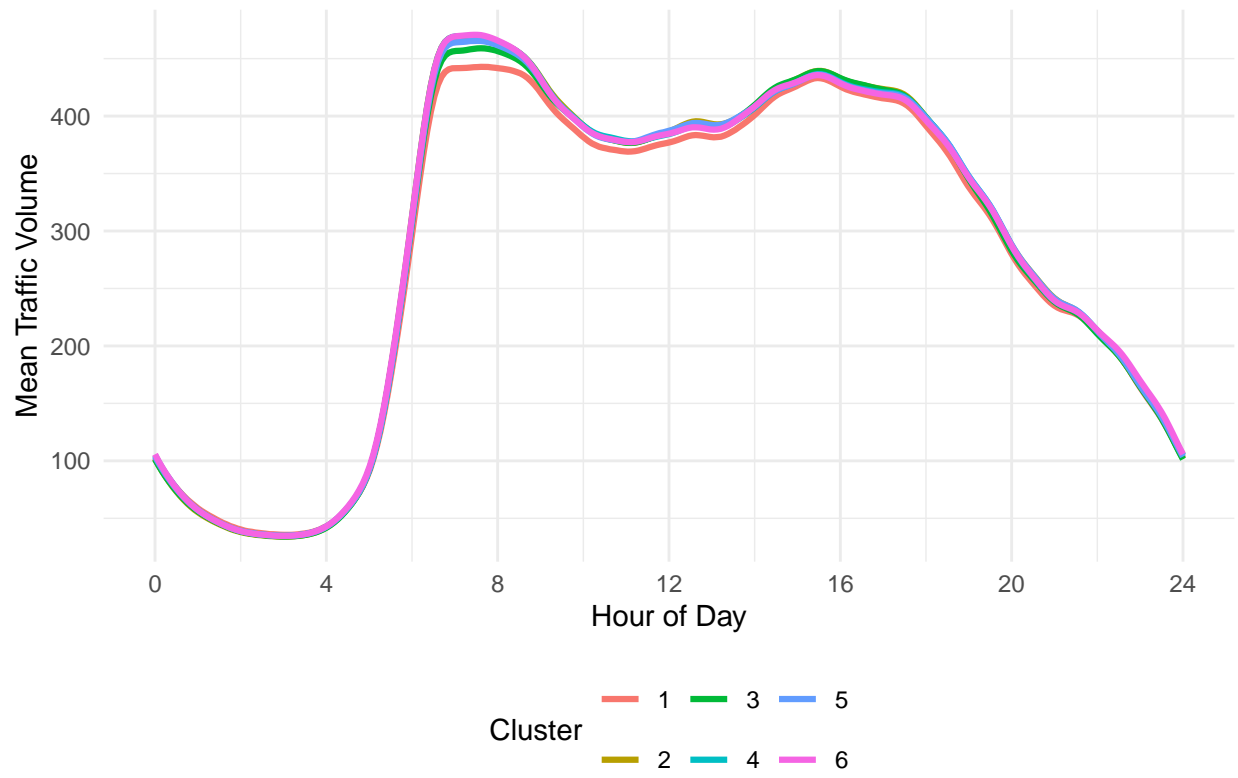




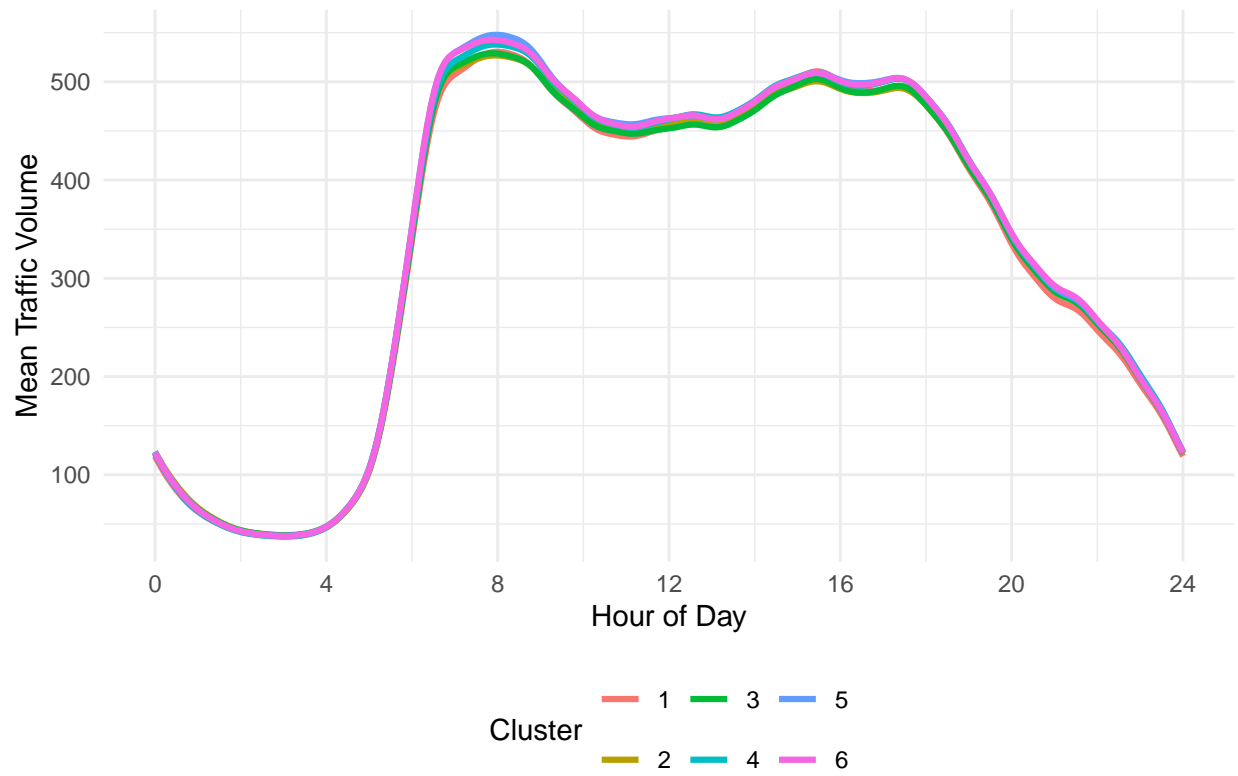
Mean Traffic Curves by Cluster – Loc15



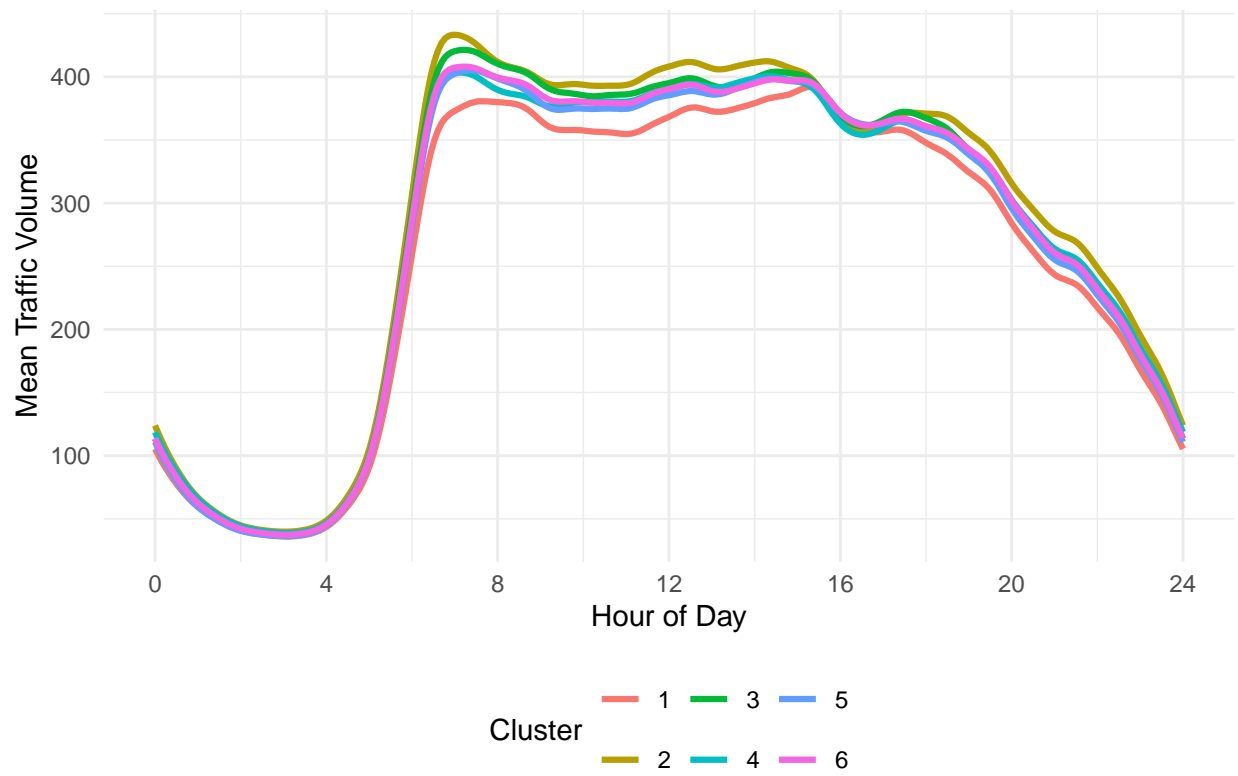
Mean Traffic Curves by Cluster – Loc16

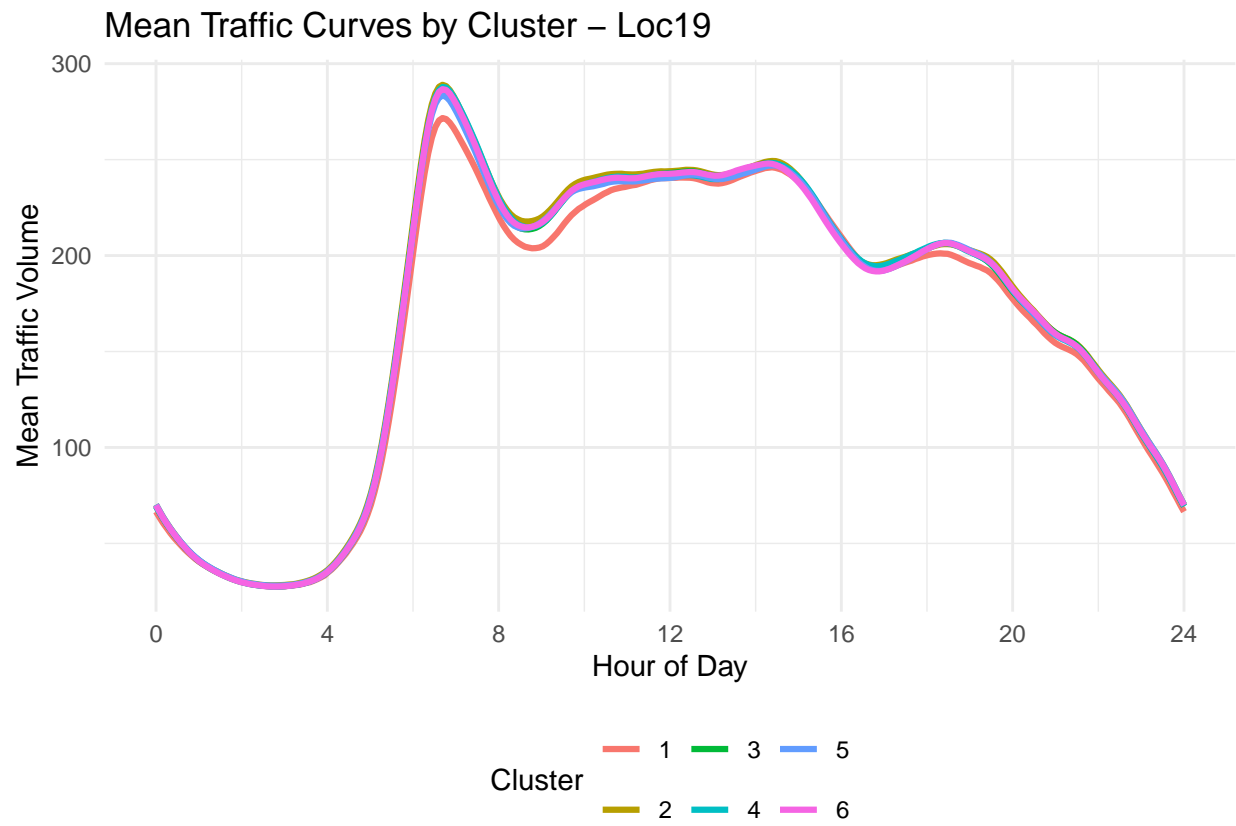


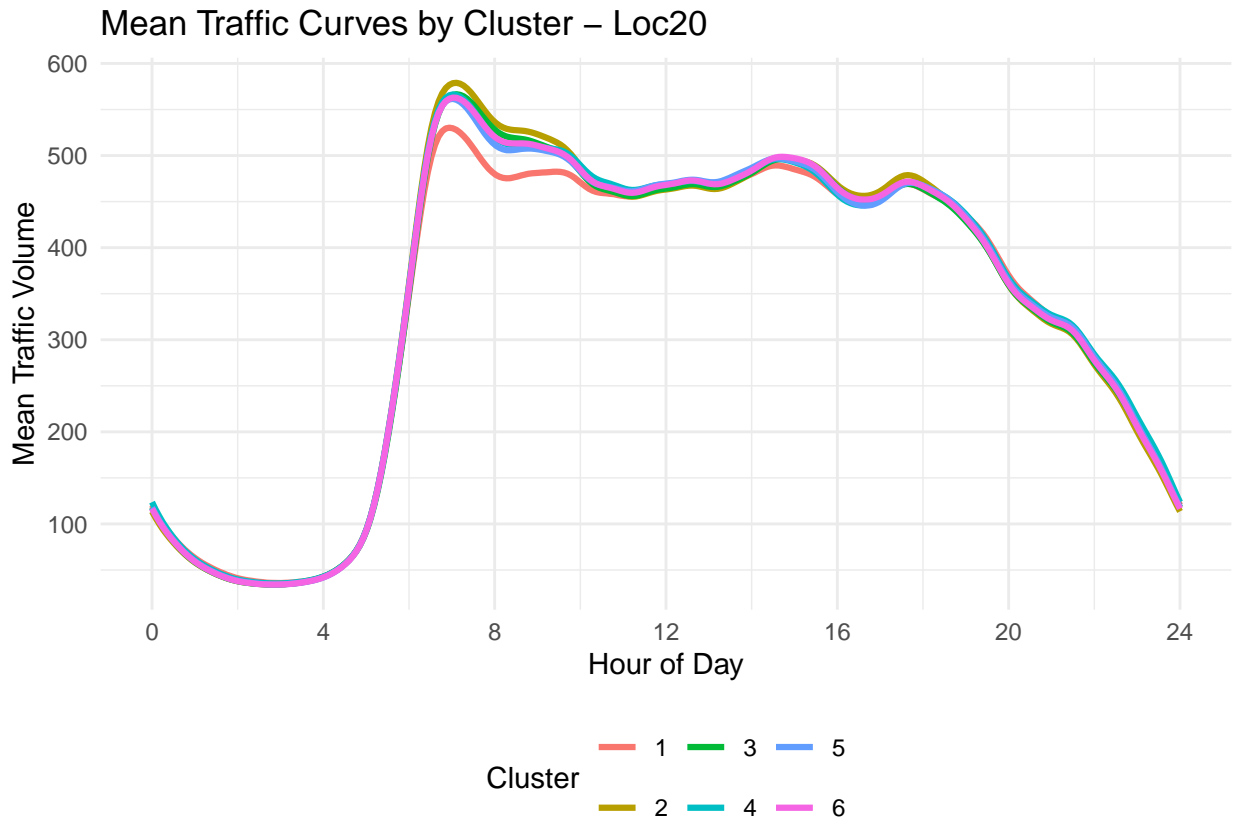
Mean Traffic Curves by Cluster – Loc17



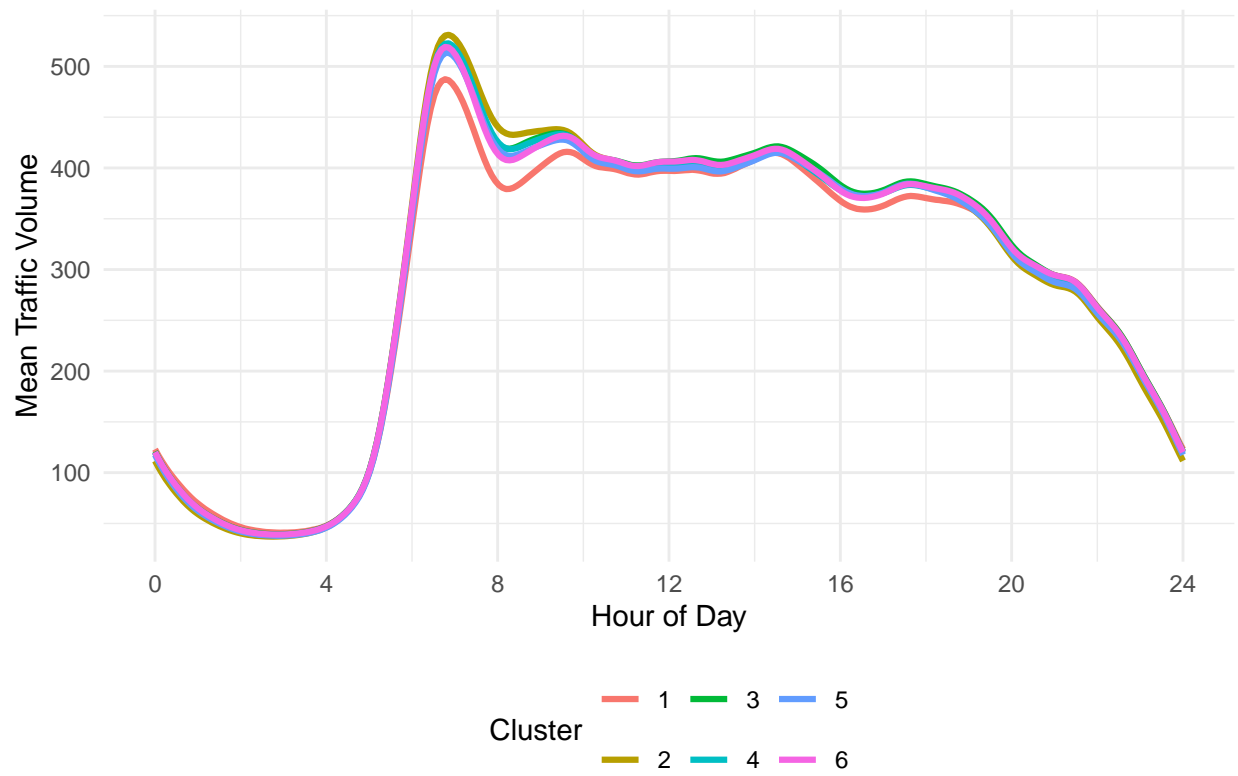
Mean Traffic Curves by Cluster – Loc18



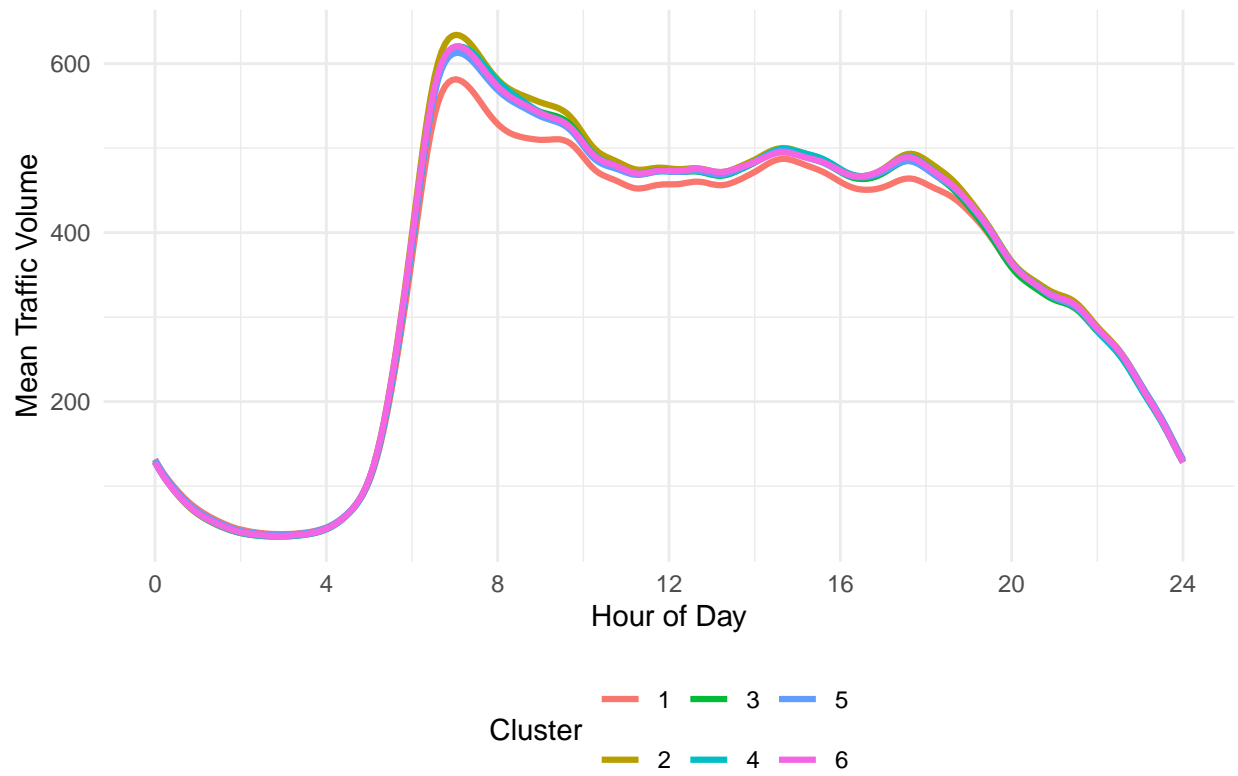




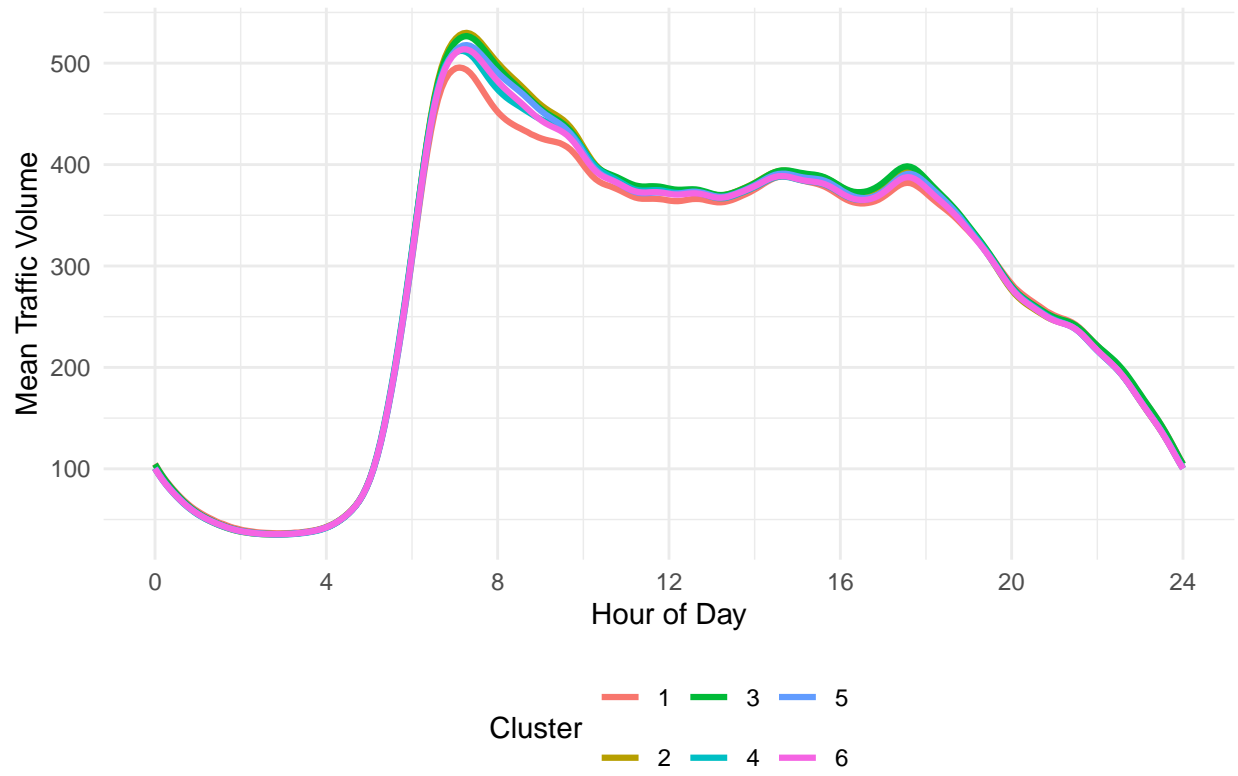
Mean Traffic Curves by Cluster – Loc21



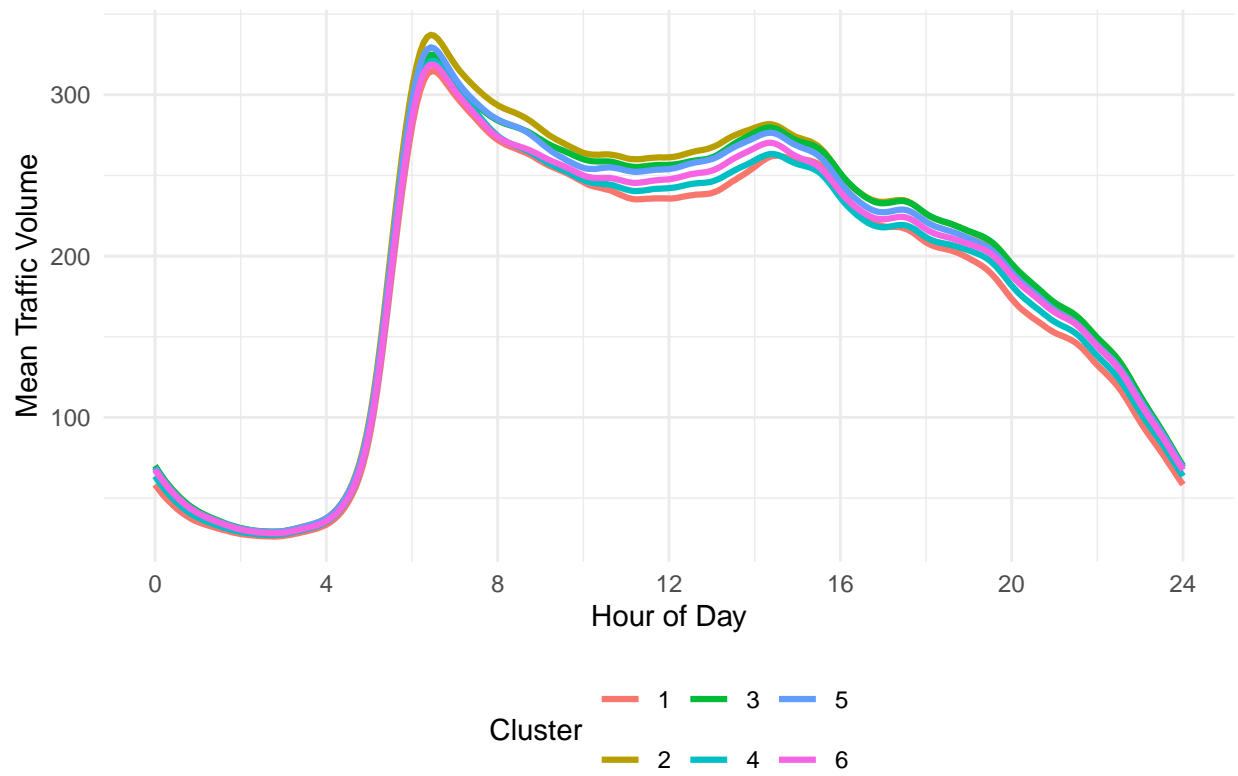
Mean Traffic Curves by Cluster – Loc22



Mean Traffic Curves by Cluster – Loc23



Mean Traffic Curves by Cluster – Loc24



Mean Traffic Curves by Cluster – Loc25

