10/29

2025-10-29

```
data22 <- read.csv("/Users/orynyehoshua/Desktop/Capstone Class/Capstone Data/updated_pitches_22.csv")
data23 <- read.csv("/Users/orynyehoshua/Desktop/Capstone Class/Capstone Data/updated_pitches_23.csv")
data24 <- read.csv("/Users/orynyehoshua/Desktop/Capstone Class/Capstone Data/updated_pitches_24.csv")

datacombined <- rbind(data22,data23,data24)

LHPLHH <- read.csv("Capstone Data/LHP_LHH_GIDP_ByPitchCluster.csv")
RHPLHH <-read.csv("Capstone Data/RHP_LHH_GIDP_ByPitchCluster.csv")
LHPRHH <- read.csv("Capstone Data/LHP_RHH_GIDP_ByPitchCluster.csv")
RHPRHH <-read.csv("Capstone Data/RHP_RHH_GIDP_ByPitchCluster.csv")

table(LHPLHH$Cluster, LHPLHH$pitchname_desc)
```

```
##
##     Changeup Curveball Cutter Four-seam FB Knuckle Curve Sinker Slider Slurve
##   0        0      1462     59            0           174      0   1683     85
##   1        8         0    122         6190             0    831     33      0
##   2       10       287   1509          159           113      9   4285     11
##   3      645         8     16           28            14    173     52      0
##   4      208         0     28          992             0   6000      5      0
##
##     Splitter Sweeper
##   0        0     819
##   1        0       0
##   2        4     300
##   3       43       5
##   4        1       0
```

```
table(LHPRHH$Cluster, LHPRHH$pitchname_desc)
```

```
##
##     Changeup Curveball Cutter Forkball Four-seam FB Knuckle Curve Sinker Slider
##   0        0      4842     72        0            0           513      1   2343
##   1       48         0    748        0        20828             0   1433     67
##   2       97       422   3923        0          478           361     10   6381
##   3    11890        41     63        5           95            45    338    107
##   4     3170         0     99        0         2342             0  10286     20
##
##     Slow Curve Slurve Splitter Sweeper
##   0         10    115        0     945
##   1          0      0        0       0
##   2          0     25        2     312
##   3          0      0      483       5
##   4          0      0        5       1
```

```r
table(RHPRHH$Cluster, RHPRHH$pitchname_desc)
```

```
##
##     Changeup Curveball Cutter Eephus Pitch Fastball Forkball Four-seam FB
##   0      315       711   9352            0        0        2          315
##   1       28         0   1980            0        0        1        35366
##   2     4590        42     48            1        3       34          106
##   3     2103         0     75            0        0        0         2749
##   4       10      6761    130            0        0        0           53
##
##     Knuckle Curve Knuckleball Screwball Sinker Slider Slow Curve Slurve
##   0           278           0         9     96  25242          1     26
##   1             0           0         0   4391    107          0      0
##   2            48          68         5    493    384          0      1
##   3             0           0         1  29221    111          0      0
##   4          1694           0         1     54   8602          5    324
##
##     Splitter Sweeper Two-seam FB
##   0       30    1786           0
##   1        3       0           0
##   2     2279      11           0
##   3      201       0           3
##   4        1    4569           0
```

```r
table(RHPLHH$Cluster, RHPLHH$pitchname_desc)
```

```
##
##     Changeup Curveball Cutter Eephus Pitch Fastball Forkball Four-seam FB
##   0      294       459   8056            0        0        0          438
##   1       41         0   2992            0        0        0        38212
##   2    13970        50     61            2        4       47          128
##   3     7116         1    145            0        0        0         2874
##   4       12      8300     67            0        0        0           47
##
##     Knuckle Curve Knuckleball Screwball Sinker Slider Slow Curve Slurve
##   0           222           0        11     35  10832          0     12
##   1             0           0         0   1607     89          0      0
##   2            19          97         2    307    237          0      1
##   3             0           0         2  13809     41          0      0
##   4          2008           0         2     63   3349          9    156
##
##     Splitter Sweeper Two-seam FB
##   0       12     617           0
##   1        3       0           0
##   2     4641       5           0
##   3      429       0           2
##   4        1    1506           0
```

```r
table(LHPLHH$Cluster)
```

```
##
##    0    1    2    3    4
## 4282 7184 6687  984 7234
```

```r
table(LHPRHH$Cluster)
```

```
##
##     0     1     2     3     4
##  8841 23124 12011 13072 15923
```

```r
table(RHPRHH$Cluster)
```

```
##
##     0     1     2     3     4
## 38163 41876  8113 34464 22204
```

```r
table(RHPLHH$Cluster)
```

```
##
##     0     1     2     3     4
## 20988 42944 19571 24419 15520
```

```r
table(LHPLHH$Cluster, LHPLHH$pitchname_desc )
```

```
##
##     Changeup Curveball Cutter Four-seam FB Knuckle Curve Sinker Slider Slurve
##   0        0      1462     59            0           174      0   1683     85
##   1        8         0    122         6190             0    831     33      0
##   2       10       287   1509          159           113      9   4285     11
##   3      645         8     16           28            14    173     52      0
##   4      208         0     28          992             0   6000      5      0
##
##     Splitter Sweeper
##   0        0     819
##   1        0       0
##   2        4     300
##   3       43       5
##   4        1       0
```

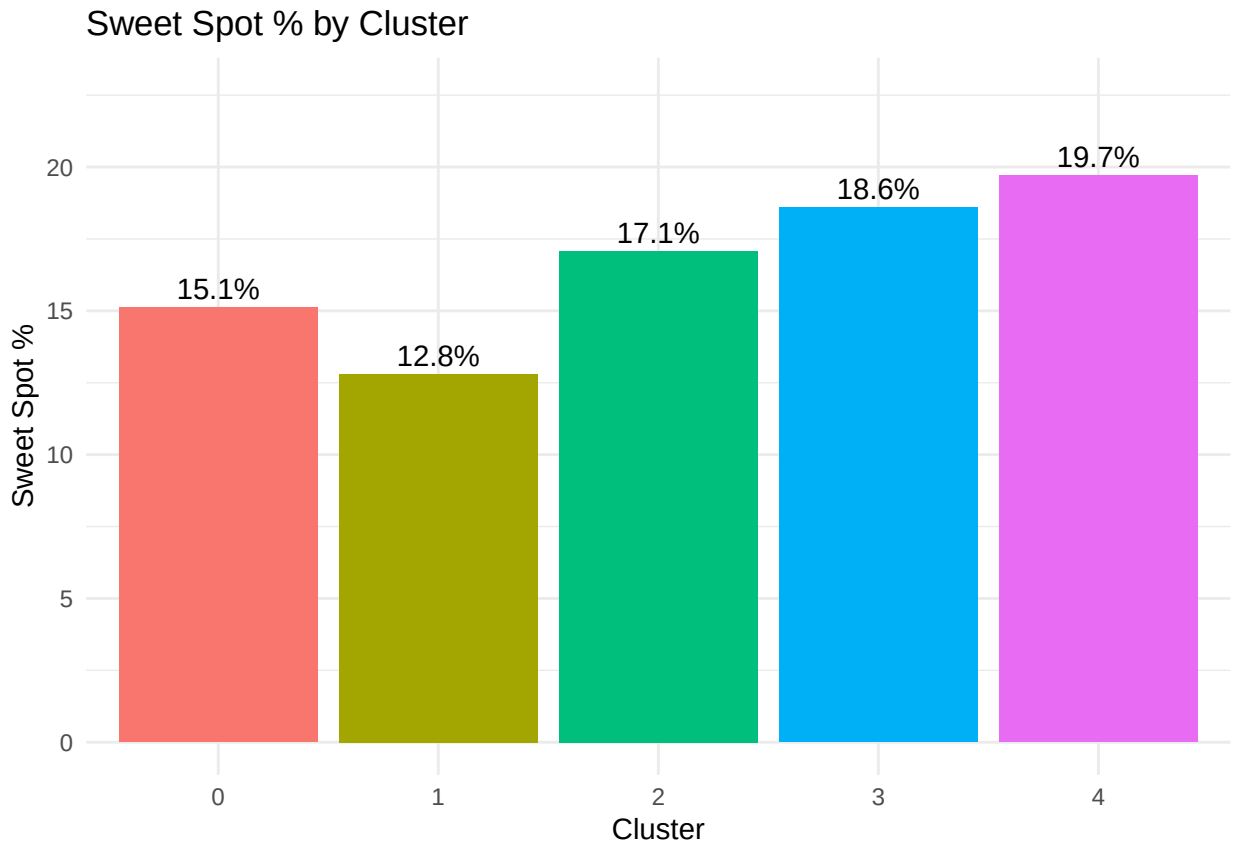
```r
library(dplyr)
```

```
##
## 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
cluster_summary <- LHPLHH %>%
  group_by(Cluster) %>%
  summarise(
    n_pitches = n(),
    usagerate = n_pitches/nrow(LHPLHH),
    n_sweetspot = sum(GIDP_SweetSpot == 1, na.rm = TRUE),
    pct_sweetspot = n_sweetspot / n_pitches * 100,
    spinrate_mean = mean(spinrate, na.rm = TRUE),
    spinrate_sd = sd(spinrate, na.rm = TRUE),
    relspeed_mean = mean(relspeed, na.rm = TRUE),
    relspeed_sd = sd(relspeed, na.rm = TRUE),
    inducedvertbreak_mean = mean(inducedvertbreak, na.rm = TRUE),
    inducedvertbreak_sd = sd(inducedvertbreak, na.rm = TRUE),
    horzbreak_mean = mean(horzbreak, na.rm = TRUE),
    horzbreak_sd = sd(horzbreak, na.rm = TRUE)
  )

# Step 2: pitch rankings (top 2 pitch types per cluster)
pitchrankings <- LHPLHH %>%
  group_by(Cluster, pitchname_desc) %>%
  summarise(count = n(), .groups = "drop_last") %>%
  mutate(pct_in_cluster = count / sum(count) * 100) %>%
  arrange(Cluster, desc(pct_in_cluster)) %>%
  group_by(Cluster) %>%
  slice_head(n = 2) %>%
  mutate(rank = row_number()) %>%
  ungroup() %>%
  tidyr::pivot_wider(
    id_cols = Cluster,
    names_from = rank,
    values_from = c(pitchname_desc, pct_in_cluster),
    names_glue = "pitch{rank}_{.value}"
  )

# Step 3: join them
cluster_summary_full <- dplyr::left_join(cluster_summary, pitchrankings, by = "Cluster")

cluster_summary_full
```

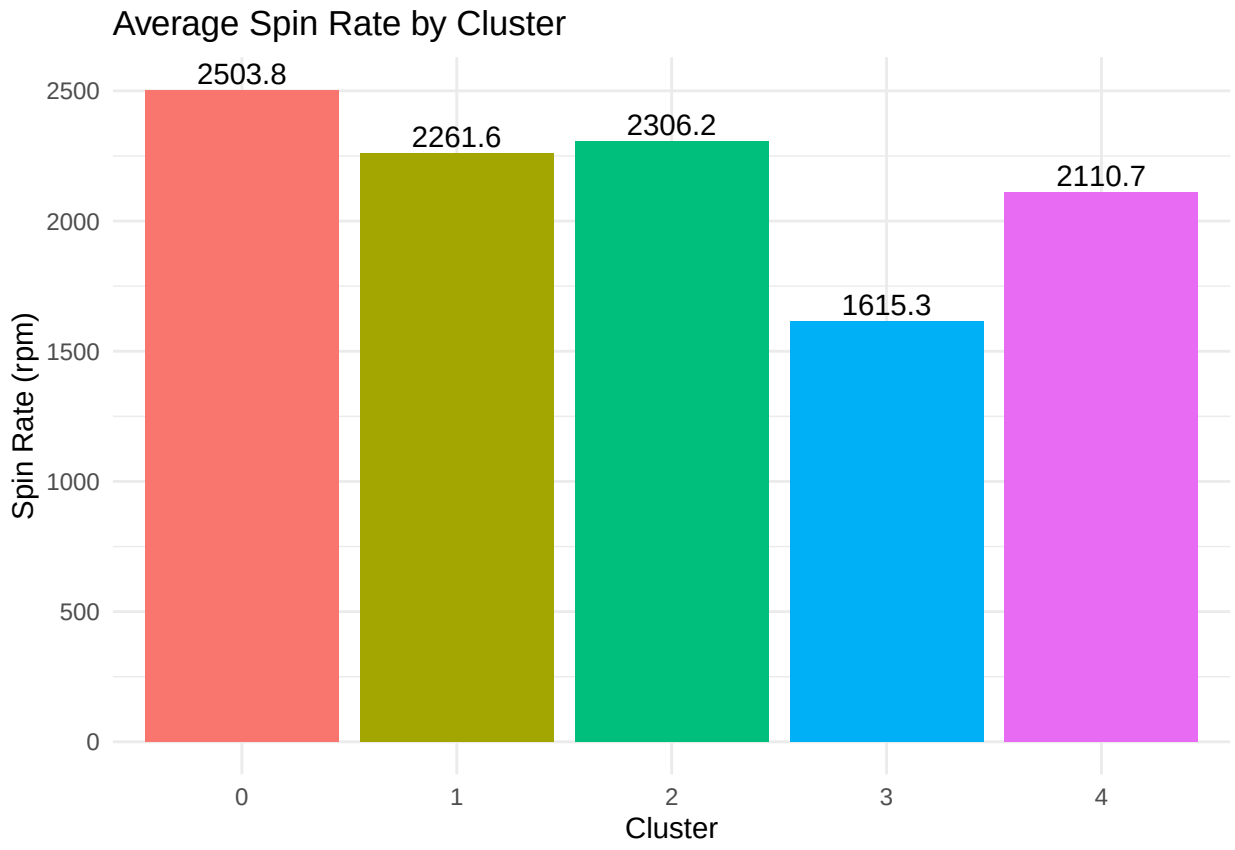
```
## # A tibble: 5 x 17
##   Cluster n_pitches usagerate n_sweetspot pct_sweetspot spinrate_mean
##     <int>     <int>     <dbl>       <int>         <dbl>         <dbl>
## 1       0      4282    0.162          647          15.1         2504.
## 2       1      7184    0.272          920          12.8         2262.
## 3       2      6687    0.254         1142          17.1         2306.
## 4       3       984    0.0373         183          18.6         1615.
## 5       4      7234    0.274         1426          19.7         2111.
## # i 11 more variables: spinrate_sd <dbl>, relspeed_mean <dbl>,
## #   relspeed_sd <dbl>, inducedvertbreak_mean <dbl>, inducedvertbreak_sd <dbl>,
## #   horzbreak_mean <dbl>, horzbreak_sd <dbl>, pitch1_pitchname_desc <chr>,
## #   pitch2_pitchname_desc <chr>, pitch1_pct_in_cluster <dbl>,
## #   pitch2_pct_in_cluster <dbl>
```

```r
library(ggplot2)
ggplot(cluster_summary_full,
       aes(x = factor(Cluster),
           y = pct_sweetspot,
           fill = factor(Cluster))) +
  geom_col(show.legend = FALSE) +
  geom_text(aes(label = sprintf("%.1f%%", pct_sweetspot)),
            vjust = -0.4,
            size = 4) +
  labs(
    title = "Sweet Spot % by Cluster",
    x = "Cluster",
    y = "Sweet Spot %"
  ) +
  theme_minimal() +
  ylim(0, max(cluster_summary_full$pct_sweetspot) * 1.15)
```



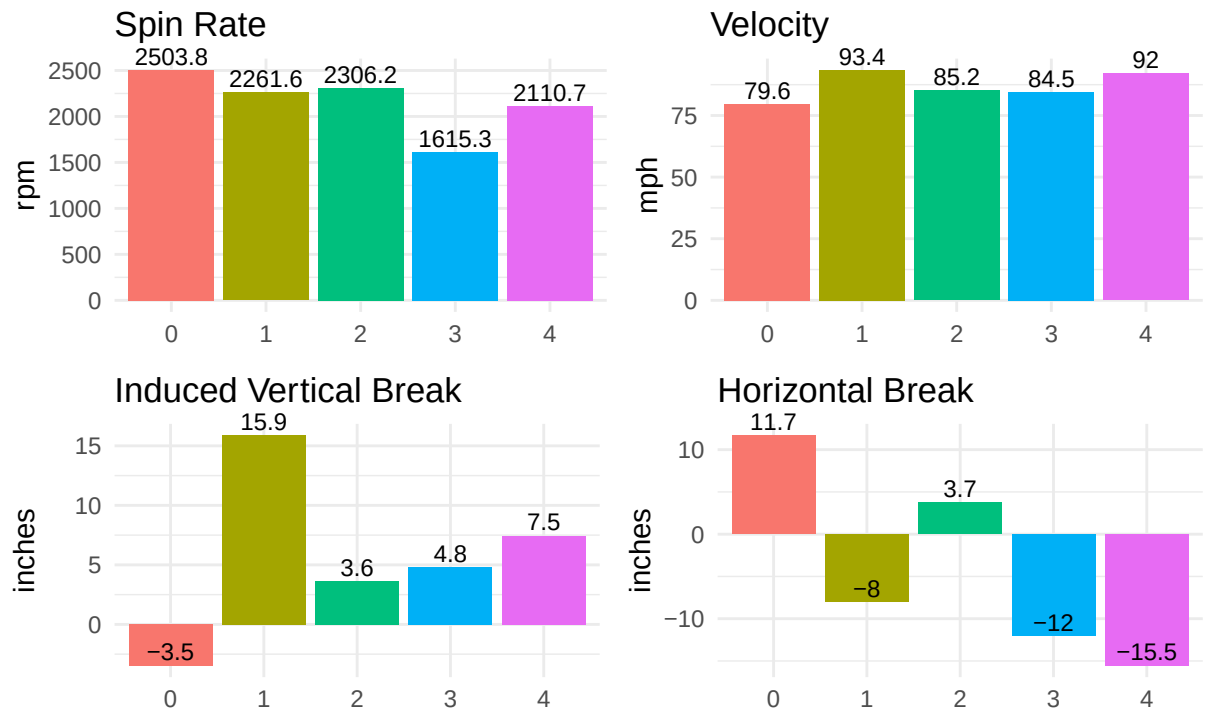

```r
library(tidyr)
cluster_metrics_long <- cluster_summary_full %>%
  select(Cluster,
         relspeed_mean,
         spinrate_mean,
         inducedvertbreak_mean,
         horzbreak_mean) %>%
  pivot_longer(
```

```
    cols = -Cluster,
    names_to = "metric",
    values_to = "value"
  )


library(ggplot2)
library(dplyr)
library(patchwork)

# Single plot (works as-is)
ggplot(cluster_metrics_long %>% filter(metric == "spinrate_mean"),
       aes(x = factor(Cluster), y = value, fill = factor(Cluster))) +
  geom_col(show.legend = FALSE) +
  geom_text(aes(label = round(value, 1)), vjust = -0.3, size = 4) +
  labs(title = "Average Spin Rate by Cluster",
       x = "Cluster", y = "Spin Rate (rpm)") +
  theme_minimal()
```




```
# Patchwork components - FIXED
p1 <- ggplot(cluster_metrics_long %>% filter(metric == "spinrate_mean"),
             aes(x = factor(Cluster), y = value, fill = factor(Cluster))) +
  geom_col(show.legend = FALSE) +
  geom_text(aes(label = round(value, 1)), vjust = -0.3, size = 3) +  # <- closed here
  labs(title = "Spin Rate", x = NULL, y = "rpm") +
```

```
  theme_minimal() +
  coord_cartesian(clip = "off")

p2 <- ggplot(cluster_metrics_long %>% filter(metric == "relspeed_mean"),
             aes(x = factor(Cluster), y = value, fill = factor(Cluster))) +
  geom_col(show.legend = FALSE) +
  geom_text(aes(label = round(value, 1)), vjust = -0.3, size = 3) +  # <- closed here
  labs(title = "Velocity", x = NULL, y = "mph") +
  theme_minimal() +
  coord_cartesian(clip = "off")

p3 <- ggplot(cluster_metrics_long %>% filter(metric == "inducedvertbreak_mean"),
             aes(x = factor(Cluster), y = value, fill = factor(Cluster))) +
  geom_col(show.legend = FALSE) +
  geom_text(aes(label = round(value, 1)), vjust = -0.3, size = 3) +
  labs(title = "Induced Vertical Break", x = NULL, y = "inches") +
  theme_minimal() +
  coord_cartesian(clip = "off")

p4 <- ggplot(cluster_metrics_long %>% filter(metric == "horzbreak_mean"),
             aes(x = factor(Cluster), y = value, fill = factor(Cluster))) +
  geom_col(show.legend = FALSE) +
  geom_text(aes(label = round(value, 1)), vjust = -0.3, size = 3) +
  labs(title = "Horizontal Break", x = NULL, y = "inches") +
  theme_minimal() +
  coord_cartesian(clip = "off")

# Combine + title
(p1 | p2) / (p3 | p4) +
  plot_annotation(
    title = "Left-Handed Pitcher vs Left-Handed Hitter",
    subtitle = "Comparing spin rate, velocity, and break characteristics across clusters",
    theme = theme(
      plot.title = element_text(size = 18, face = "bold", hjust = 0.5),
      plot.subtitle = element_text(size = 12, hjust = 0.5, color = "gray30")
    )
  )
```

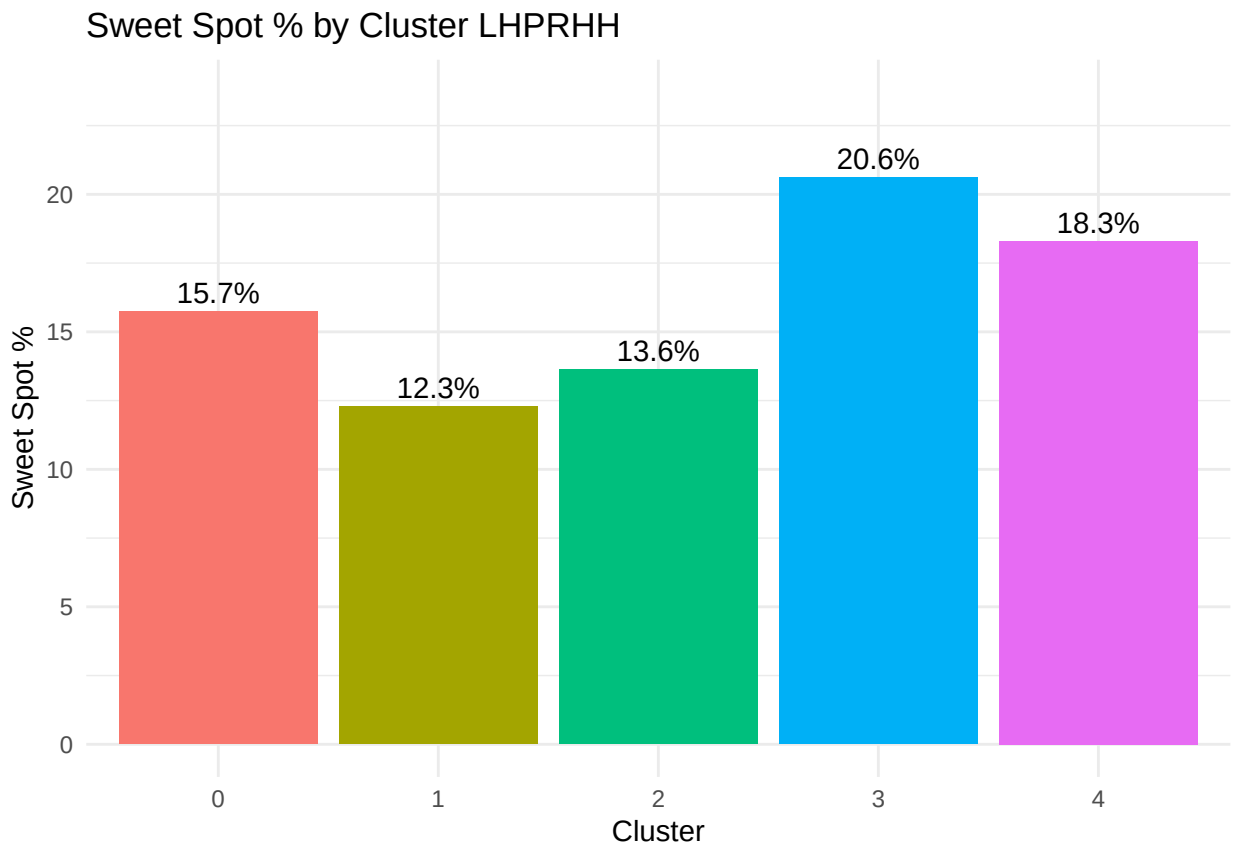



```r
library(dplyr)

cluster_summary_LHPRHH <- LHPRHH %>%
  group_by(Cluster) %>%
  summarise(
    n_pitches = n(),
    usagerate = n_pitches/nrow(LHPRHH),
    n_sweetspot = sum(GIDP_SweetSpot == 1, na.rm = TRUE),
    pct_sweetspot = n_sweetspot / n_pitches * 100,
    spinrate_mean = mean(spinrate, na.rm = TRUE),
    spinrate_sd = sd(spinrate, na.rm = TRUE),
    relspeed_mean = mean(relspeed, na.rm = TRUE),
    relspeed_sd = sd(relspeed, na.rm = TRUE),
    inducedvertbreak_mean = mean(inducedvertbreak, na.rm = TRUE),
    inducedvertbreak_sd = sd(inducedvertbreak, na.rm = TRUE),
    horzbreak_mean = mean(horzbreak, na.rm = TRUE),
    horzbreak_sd = sd(horzbreak, na.rm = TRUE)
  )

# Step 2: pitch rankings (top 2 pitch types per cluster)
pitchrankings_LHPRHH <- LHPRHH %>%
  group_by(Cluster, pitchname_desc) %>%
  summarise(count = n(), .groups = "drop_last") %>%
  mutate(pct_in_cluster = count / sum(count) * 100) %>%
  arrange(Cluster, desc(pct_in_cluster)) %>%
  group_by(Cluster) %>%
```

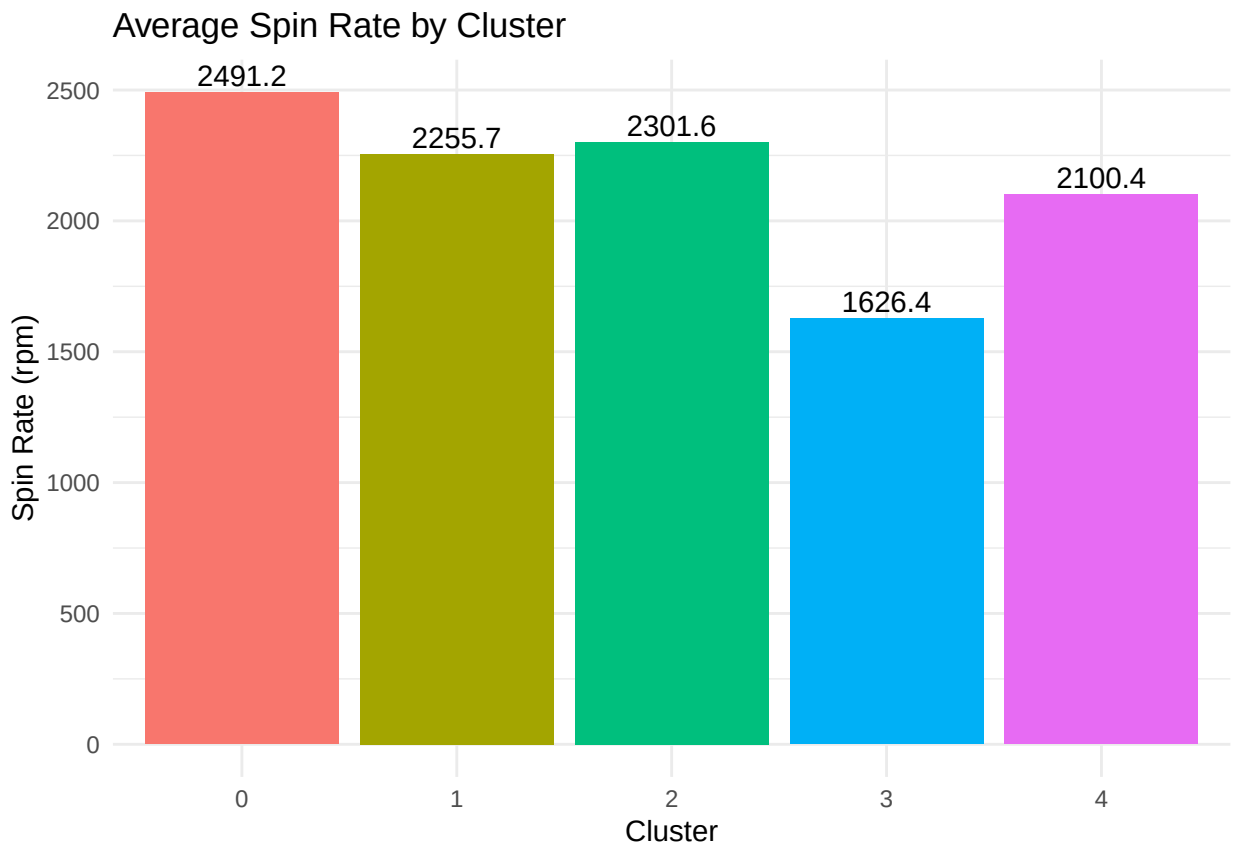
```
  slice_head(n = 2) %>%
  mutate(rank = row_number()) %>%
  ungroup() %>%
  tidyr::pivot_wider(
    id_cols = Cluster,
    names_from = rank,
    values_from = c(pitchname_desc, pct_in_cluster),
    names_glue = "pitch{rank}_{.value}"
  )

# Step 3: join them
cluster_summary_full_LHPRHH <- dplyr::left_join(cluster_summary_LHPRHH, pitchrankings_LHPRHH, by = "Clus

cluster_summary_full_LHPRHH
```

```
## # A tibble: 5 x 17
##   Cluster n_pitches usagerate n_sweetspot pct_sweetspot spinrate_mean
##     <int>     <int>     <dbl>       <int>         <dbl>         <dbl>
## 1       0      8841     0.121        1391          15.7         2491.
## 2       1     23124     0.317        2842          12.3         2256.
## 3       2     12011     0.165        1637          13.6         2302.
## 4       3     13072     0.179        2695          20.6         1626.
## 5       4     15923     0.218        2913          18.3         2100.
## # i 11 more variables: spinrate_sd <dbl>, relspeed_mean <dbl>,
## #   relspeed_sd <dbl>, inducedvertbreak_mean <dbl>, inducedvertbreak_sd <dbl>,
## #   horzbreak_mean <dbl>, horzbreak_sd <dbl>, pitch1_pitchname_desc <chr>,
## #   pitch2_pitchname_desc <chr>, pitch1_pct_in_cluster <dbl>,
## #   pitch2_pct_in_cluster <dbl>
```

```
library(ggplot2)
ggplot(cluster_summary_full_LHPRHH,
       aes(x = factor(Cluster),
           y = pct_sweetspot,
           fill = factor(Cluster))) +
  geom_col(show.legend = FALSE) +
  geom_text(aes(label = sprintf("%.1f%%", pct_sweetspot)),
            vjust = -0.4,
            size = 4) +
  labs(
    title = "Sweet Spot % by Cluster LHPRHH",
    x = "Cluster",
    y = "Sweet Spot %"
  ) +
  theme_minimal() +
  ylim(0, max(cluster_summary_full_LHPRHH$pct_sweetspot) * 1.15)
```

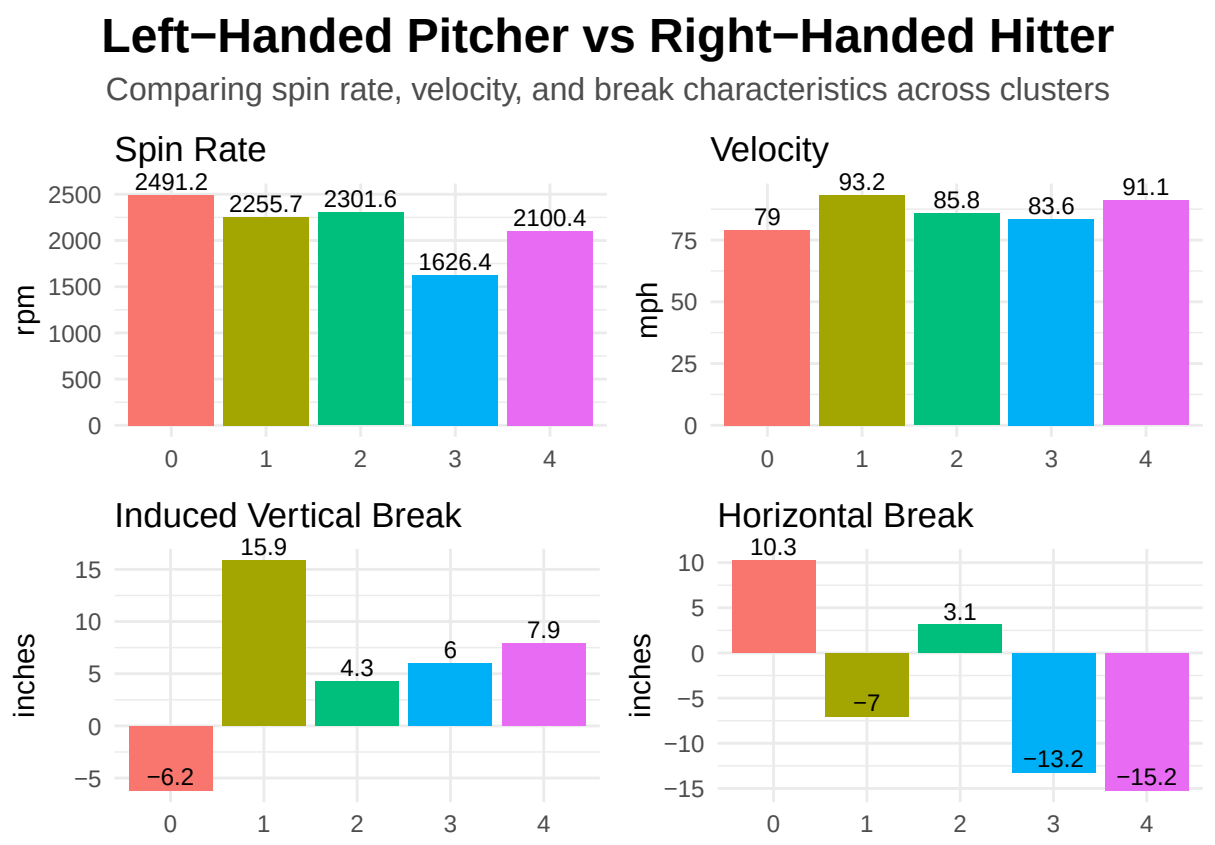



```
cluster_metrics_long_LHPRHH <- cluster_summary_full_LHPRHH %>%
  select(Cluster, spinrate_mean, relspeed_mean, inducedvertbreak_mean, horzbreak_mean) %>%
  pivot_longer(
    cols = -Cluster,
    names_to = "metric",
    values_to = "value"
  )

ggplot(cluster_metrics_long_LHPRHH %>% filter(metric == "spinrate_mean"),
       aes(x = factor(Cluster), y = value, fill = factor(Cluster))) +
  geom_col(show.legend = FALSE) +
  geom_text(aes(label = round(value, 1)), vjust = -0.3, size = 4) +
  labs(
    title = "Average Spin Rate by Cluster",
    x = "Cluster",
    y = "Spin Rate (rpm)"
  ) +
  theme_minimal()
```




```
library(patchwork)

p1LR <- ggplot(cluster_metrics_long_LHPRHH %>% filter(metric == "spinrate_mean"),
               aes(x = factor(Cluster), y = value, fill = factor(Cluster))) +
  geom_col(show.legend = FALSE) +
  geom_text(aes(label = round(value, 1)), vjust = -0.3, size = 3) +   # smaller labels
  labs(title = "Spin Rate", x = NULL, y = "rpm") +
  theme_minimal() +
  coord_cartesian(clip = "off")

p2LR <- ggplot(cluster_metrics_long_LHPRHH %>% filter(metric == "relspeed_mean"),
               aes(x = factor(Cluster), y = value, fill = factor(Cluster))) +
  geom_col(show.legend = FALSE) +
  geom_text(aes(label = round(value, 1)), vjust = -0.3, size = 3) +   # smaller labels
  labs(title = "Velocity", x = NULL, y = "mph") +
  theme_minimal() +
  coord_cartesian(clip = "off")

p3LR <- ggplot(cluster_metrics_long_LHPRHH %>% filter(metric == "inducedvertbreak_mean"),
               aes(x = factor(Cluster), y = value, fill = factor(Cluster))) +
  geom_col(show.legend = FALSE) +
  geom_text(aes(label = round(value, 1)), vjust = -0.3, size = 3) +   # smaller labels
  labs(title = "Induced Vertical Break", x = NULL, y = "inches") +
  theme_minimal() +
  coord_cartesian(clip = "off")
```

```r
p4LR <- ggplot(cluster_metrics_long_LHPRHH %>% filter(metric == "horzbreak_mean"),
               aes(x = factor(Cluster), y = value, fill = factor(Cluster))) +
  geom_col(show.legend = FALSE) +
  geom_text(aes(label = round(value, 1)), vjust = -0.3, size = 3) +   # smaller labels
  labs(title = "Horizontal Break", x = NULL, y = "inches") +
  theme_minimal() +
  coord_cartesian(clip = "off")

# Combine in a 2x2 grid with title/subtitle
(p1LR | p2LR) / (p3LR | p4LR) +
  plot_annotation(
    title = "Left-Handed Pitcher vs Right-Handed Hitter",
    subtitle = "Comparing spin rate, velocity, and break characteristics across clusters",
    theme = theme(
      plot.title = element_text(size = 18, face = "bold", hjust = 0.5),
      plot.subtitle = element_text(size = 12, hjust = 0.5, color = "gray30")
    )
  )
```



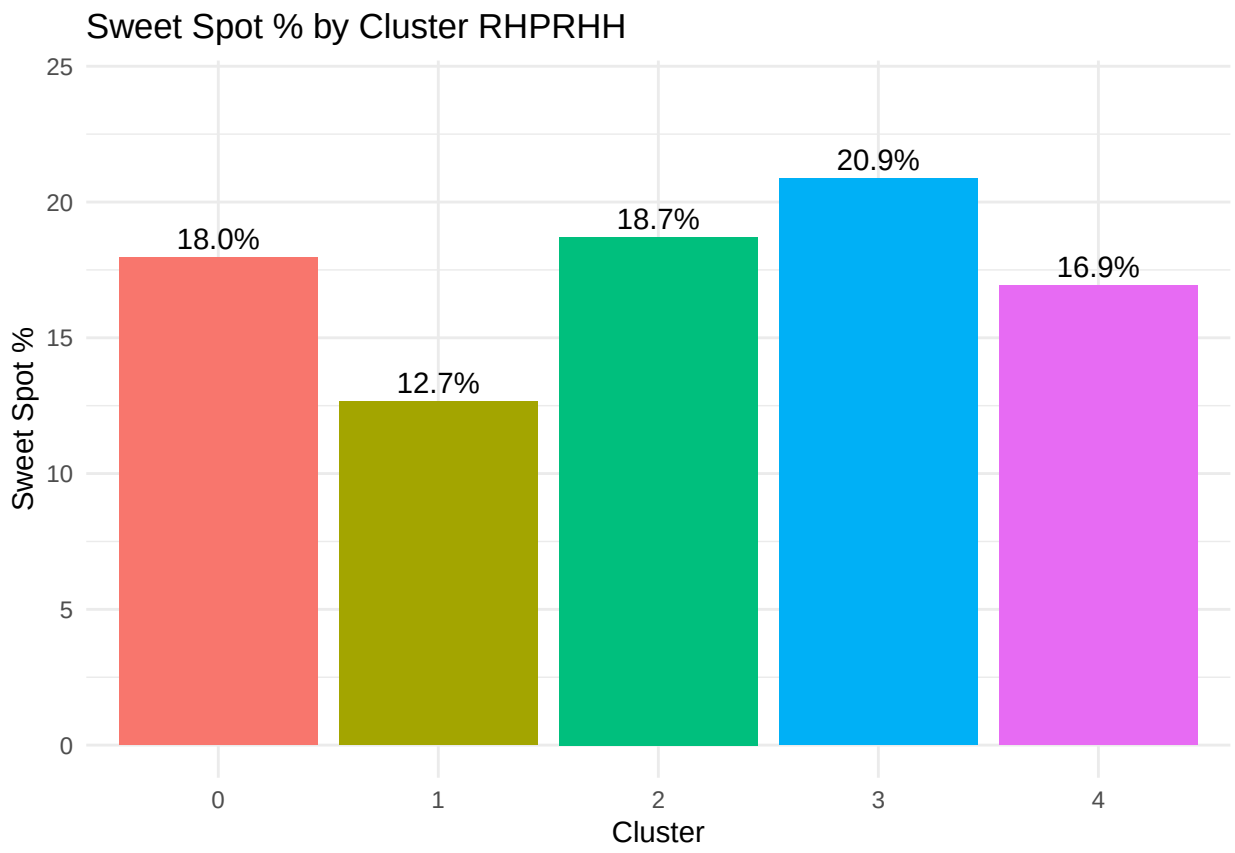

```r
library(dplyr)

cluster_summary_RHPRHH <- RHPRHH %>%
  group_by(Cluster) %>%
  summarise(
    n_pitches = n(),
```

```
    usagerate = n_pitches/nrow(RHPRHH),
    n_sweetspot = sum(GIDP_SweetSpot == 1, na.rm = TRUE),
    pct_sweetspot = n_sweetspot / n_pitches * 100,
    spinrate_mean = mean(spinrate, na.rm = TRUE),
    spinrate_sd = sd(spinrate, na.rm = TRUE),
    relspeed_mean = mean(relspeed, na.rm = TRUE),
    relspeed_sd = sd(relspeed, na.rm = TRUE),
    inducedvertbreak_mean = mean(inducedvertbreak, na.rm = TRUE),
    inducedvertbreak_sd = sd(inducedvertbreak, na.rm = TRUE),
    horzbreak_mean = mean(horzbreak, na.rm = TRUE),
    horzbreak_sd = sd(horzbreak, na.rm = TRUE)
  )

# Step 2: pitch rankings (top 2 pitch types per cluster)
pitchrankings_RHPRHH <- RHPRHH %>%
  group_by(Cluster, pitchname_desc) %>%
  summarise(count = n(), .groups = "drop_last") %>%
  mutate(pct_in_cluster = count / sum(count) * 100) %>%
  arrange(Cluster, desc(pct_in_cluster)) %>%
  group_by(Cluster) %>%
  slice_head(n = 2) %>%
  mutate(rank = row_number()) %>%
  ungroup() %>%
  tidyr::pivot_wider(
    id_cols = Cluster,
    names_from = rank,
    values_from = c(pitchname_desc, pct_in_cluster),
    names_glue = "pitch{rank}_{.value}"
  )

# Step 3: join them
cluster_summary_full_RHPRHH <- dplyr::left_join(cluster_summary_RHPRHH, pitchrankings_RHPRHH, by = "Clus

cluster_summary_full_RHPRHH
```

```
## # A tibble: 5 x 17
##   Cluster n_pitches usagerate n_sweetspot pct_sweetspot spinrate_mean
##     <int>     <int>     <dbl>       <int>         <dbl>         <dbl>
## 1       0     38163    0.264         6851          18.0         2393.
## 2       1     41876    0.289         5303          12.7         2311.
## 3       2      8113    0.0560        1518          18.7         1518.
## 4       3     34464    0.238         7195          20.9         2137.
## 5       4     22204    0.153         3757          16.9         2602.
## # i 11 more variables: spinrate_sd <dbl>, relspeed_mean <dbl>,
## #   relspeed_sd <dbl>, inducedvertbreak_mean <dbl>, inducedvertbreak_sd <dbl>,
## #   horzbreak_mean <dbl>, horzbreak_sd <dbl>, pitch1_pitchname_desc <chr>,
## #   pitch2_pitchname_desc <chr>, pitch1_pct_in_cluster <dbl>,
## #   pitch2_pct_in_cluster <dbl>
```

```
library(patchwork)
library(ggplot2)
ggplot(cluster_summary_full_RHPRHH,
       aes(x = factor(Cluster),
```

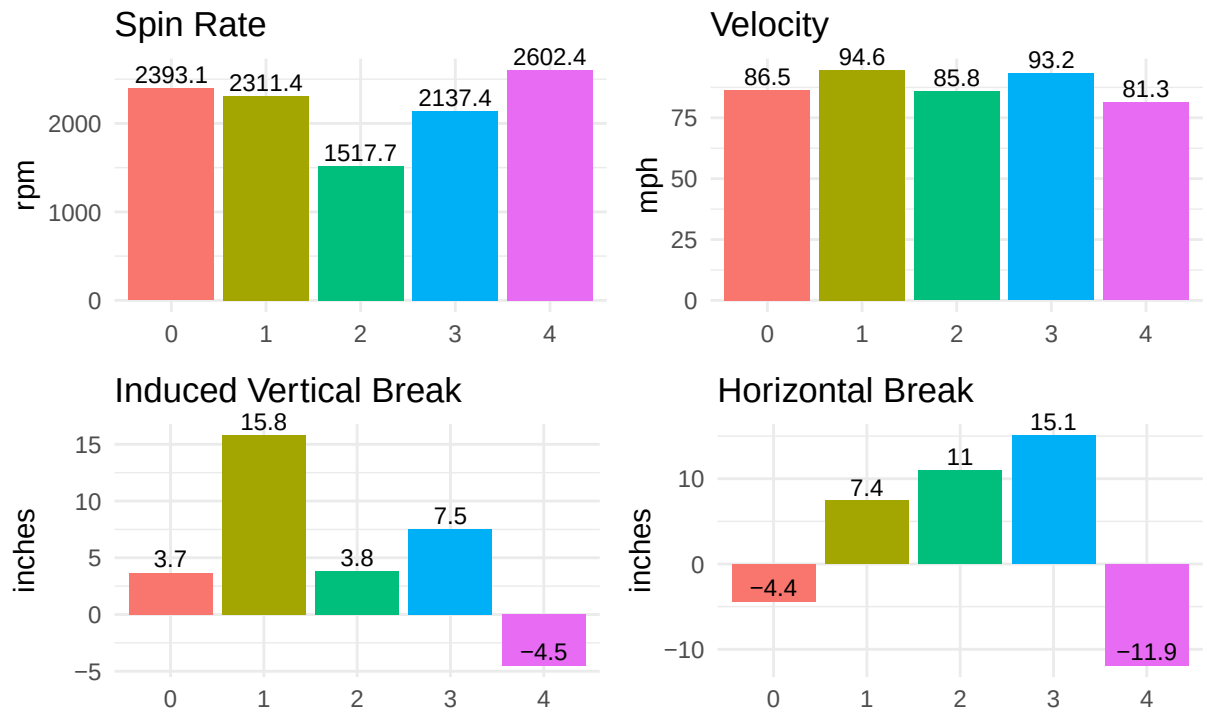
```
          y = pct_sweetspot,
          fill = factor(Cluster))) +
  geom_col(show.legend = FALSE) +
  geom_text(aes(label = sprintf("%.1f%%", pct_sweetspot)),
            vjust = -0.4,
            size = 4) +
  labs(
    title = "Sweet Spot % by Cluster RHPRHH",
    x = "Cluster",
    y = "Sweet Spot %"
  ) +
  theme_minimal() +
  ylim(0, max(cluster_summary_full_RHPRHH$pct_sweetspot) * 1.15)
```




```
cluster_metrics_long_RHPRHH <- cluster_summary_full_RHPRHH %>%
  dplyr::select(Cluster, spinrate_mean, relspeed_mean, inducedvertbreak_mean, horzbreak_mean) %>%
  tidyr::pivot_longer(
    cols = -Cluster,
    names_to = "metric",
    values_to = "value"
  )


library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v forcats   1.0.0     v readr     2.1.5
## v lubridate 1.9.4     v stringr   1.5.1
## v purrr     1.1.0     v tibble    3.3.0
## -- 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 to become errors
```

```r
library(patchwork)

p1RR <- ggplot(cluster_metrics_long_RHPRHH %>% filter(metric == "spinrate_mean"),
               aes(x = factor(Cluster), y = value, fill = factor(Cluster))) +
  geom_col(show.legend = FALSE) +
  geom_text(aes(label = round(value, 1)), vjust = -0.3, size = 3) +
  labs(title = "Spin Rate", x = NULL, y = "rpm") +
  theme_minimal() +
  coord_cartesian(clip = "off")

p2RR <- ggplot(cluster_metrics_long_RHPRHH %>% filter(metric == "relspeed_mean"),
               aes(x = factor(Cluster), y = value, fill = factor(Cluster))) +
  geom_col(show.legend = FALSE) +
  geom_text(aes(label = round(value, 1)), vjust = -0.3, size = 3) +
  labs(title = "Velocity", x = NULL, y = "mph") +
  theme_minimal() +
  coord_cartesian(clip = "off")

p3RR <- ggplot(cluster_metrics_long_RHPRHH %>% filter(metric == "inducedvertbreak_mean"),
               aes(x = factor(Cluster), y = value, fill = factor(Cluster))) +
  geom_col(show.legend = FALSE) +
  geom_text(aes(label = round(value, 1)), vjust = -0.3, size = 3) +
  labs(title = "Induced Vertical Break", x = NULL, y = "inches") +
  theme_minimal() +
  coord_cartesian(clip = "off")

p4RR <- ggplot(cluster_metrics_long_RHPRHH %>% filter(metric == "horzbreak_mean"),
               aes(x = factor(Cluster), y = value, fill = factor(Cluster))) +
  geom_col(show.legend = FALSE) +
  geom_text(aes(label = round(value, 1)), vjust = -0.3, size = 3) +
  labs(title = "Horizontal Break", x = NULL, y = "inches") +
  theme_minimal() +
  coord_cartesian(clip = "off")

((p1RR | p2RR) / (p3RR | p4RR)) +
  plot_annotation(
    title = "Right-Handed Pitcher vs Right-Handed Hitter",
    subtitle = "Spin rate, velocity, and movement by cluster",
    theme = theme(
      plot.title = element_text(size = 18, face = "bold", hjust = 0.5),
      plot.subtitle = element_text(size = 12, hjust = 0.5, color = "gray30")
    )
  )
```

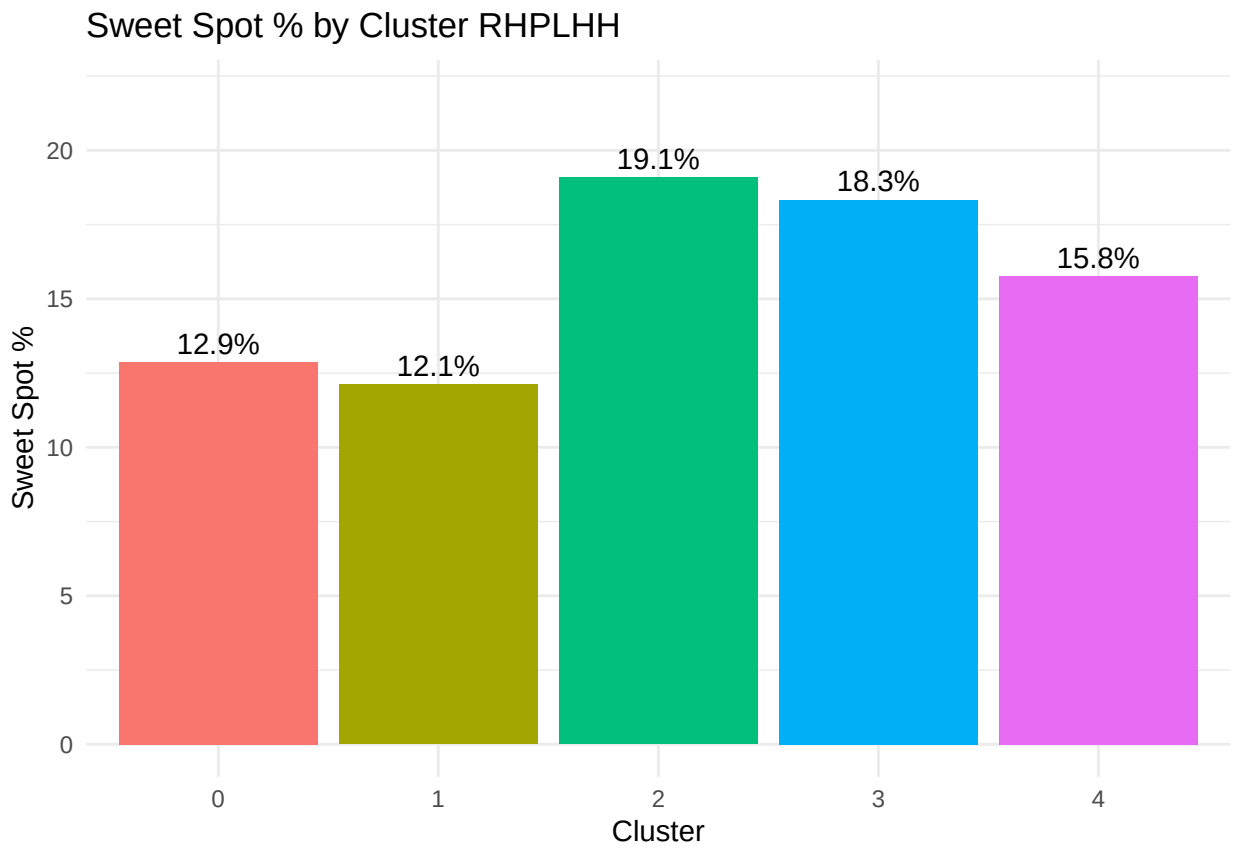



```r
library(dplyr)

cluster_summary_RHPLHH <- RHPLHH %>%
  group_by(Cluster) %>%
  summarise(
    n_pitches = n(),
    usagerate = n_pitches/nrow(RHPLHH),
    n_sweetspot = sum(GIDP_SweetSpot == 1, na.rm = TRUE),
    pct_sweetspot = n_sweetspot / n_pitches * 100,
    spinrate_mean = mean(spinrate, na.rm = TRUE),
    spinrate_sd = sd(spinrate, na.rm = TRUE),
    relspeed_mean = mean(relspeed, na.rm = TRUE),
    relspeed_sd = sd(relspeed, na.rm = TRUE),
    inducedvertbreak_mean = mean(inducedvertbreak, na.rm = TRUE),
    inducedvertbreak_sd = sd(inducedvertbreak, na.rm = TRUE),
    horzbreak_mean = mean(horzbreak, na.rm = TRUE),
    horzbreak_sd = sd(horzbreak, na.rm = TRUE)
  )

# Step 2: pitch rankings (top 2 pitch types per cluster)
pitchrankings_RHPLHH <- RHPLHH %>%
  group_by(Cluster, pitchname_desc) %>%
  summarise(count = n(), .groups = "drop_last") %>%
  mutate(pct_in_cluster = count / sum(count) * 100) %>%
  arrange(Cluster, desc(pct_in_cluster)) %>%
  group_by(Cluster) %>%
```

```
  slice_head(n = 2) %>%
  mutate(rank = row_number()) %>%
  ungroup() %>%
  tidyr::pivot_wider(
    id_cols = Cluster,
    names_from = rank,
    values_from = c(pitchname_desc, pct_in_cluster),
    names_glue = "pitch{rank}_{.value}"
  )

# Step 3: join them
cluster_summary_full_RHPLHH <- dplyr::left_join(cluster_summary_RHPLHH, pitchrankings_RHPLHH, by = "Clus

cluster_summary_full_RHPLHH
```

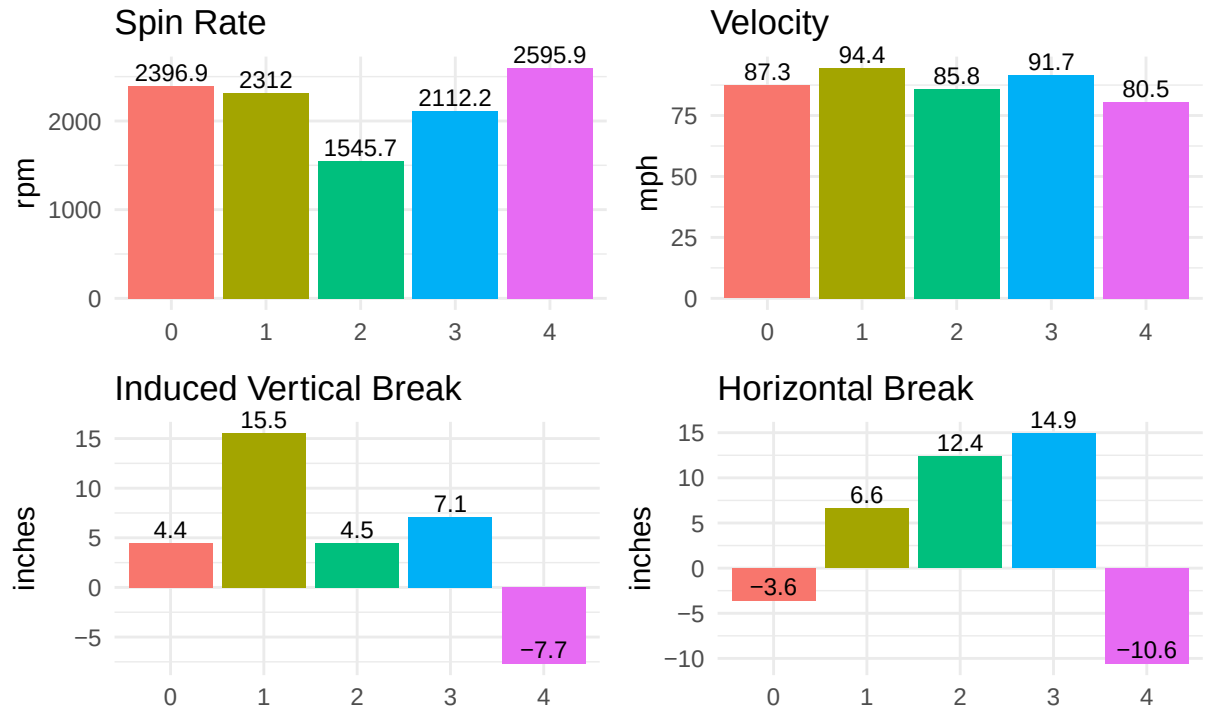
```
## # A tibble: 5 x 17
##   Cluster n_pitches usagerate n_sweetspot pct_sweetspot spinrate_mean
##     <int>     <int>     <dbl>       <int>         <dbl>         <dbl>
## 1       0     20988     0.170        2701          12.9         2397.
## 2       1     42944     0.348        5203          12.1         2312.
## 3       2     19571     0.159        3736          19.1         1546.
## 4       3     24419     0.198        4477          18.3         2112.
## 5       4     15520     0.126        2447          15.8         2596.
## # i 11 more variables: spinrate_sd <dbl>, relspeed_mean <dbl>,
## #   relspeed_sd <dbl>, inducedvertbreak_mean <dbl>, inducedvertbreak_sd <dbl>,
## #   horzbreak_mean <dbl>, horzbreak_sd <dbl>, pitch1_pitchname_desc <chr>,
## #   pitch2_pitchname_desc <chr>, pitch1_pct_in_cluster <dbl>,
## #   pitch2_pct_in_cluster <dbl>
```

```
library(ggplot2)
ggplot(cluster_summary_full_RHPLHH,
       aes(x = factor(Cluster),
           y = pct_sweetspot,
           fill = factor(Cluster))) +
  geom_col(show.legend = FALSE) +
  geom_text(aes(label = sprintf("%.1f%%", pct_sweetspot)),
            vjust = -0.4,
            size = 4) +
  labs(
    title = "Sweet Spot % by Cluster RHPLHH",
    x = "Cluster",
    y = "Sweet Spot %"
  ) +
  theme_minimal() +
  ylim(0, max(cluster_summary_full_RHPLHH$pct_sweetspot) * 1.15)
```




```
cluster_metrics_long_RHPLHH <- cluster_summary_full_RHPLHH %>%
  dplyr::select(Cluster, spinrate_mean, relspeed_mean, inducedvertbreak_mean, horzbreak_mean) %>%
  tidyr::pivot_longer(
    cols = -Cluster,
    names_to = "metric",
    values_to = "value"
  )


library(tidyverse)
library(patchwork)

p1RL <- ggplot(cluster_metrics_long_RHPLHH %>% filter(metric == "spinrate_mean"),
               aes(x = factor(Cluster), y = value, fill = factor(Cluster))) +
  geom_col(show.legend = FALSE) +
  geom_text(aes(label = round(value, 1)), vjust = -0.3, size = 3) +
  labs(title = "Spin Rate", x = NULL, y = "rpm") +
  theme_minimal() +
  coord_cartesian(clip = "off")

p2RL <- ggplot(cluster_metrics_long_RHPLHH %>% filter(metric == "relspeed_mean"),
               aes(x = factor(Cluster), y = value, fill = factor(Cluster))) +
  geom_col(show.legend = FALSE) +
  geom_text(aes(label = round(value, 1)), vjust = -0.3, size = 3) +
  labs(title = "Velocity", x = NULL, y = "mph") +
  theme_minimal() +
```

```
  coord_cartesian(clip = "off")

p3RL <- ggplot(cluster_metrics_long_RHPLHH %>% filter(metric == "inducedvertbreak_mean"),
               aes(x = factor(Cluster), y = value, fill = factor(Cluster))) +
  geom_col(show.legend = FALSE) +
  geom_text(aes(label = round(value, 1)), vjust = -0.3, size = 3) +
  labs(title = "Induced Vertical Break", x = NULL, y = "inches") +
  theme_minimal() +
  coord_cartesian(clip = "off")

p4RL <- ggplot(cluster_metrics_long_RHPLHH %>% filter(metric == "horzbreak_mean"),
               aes(x = factor(Cluster), y = value, fill = factor(Cluster))) +
  geom_col(show.legend = FALSE) +
  geom_text(aes(label = round(value, 1)), vjust = -0.3, size = 3) +
  labs(title = "Horizontal Break", x = NULL, y = "inches") +
  theme_minimal() +
  coord_cartesian(clip = "off")

# Combine plots in 2x2 grid with overall title/subtitle
(p1RL | p2RL) / (p3RL | p4RL) +
  plot_annotation(
    title = "Right-Handed Pitcher vs Left-Handed Hitter",
    subtitle = "Comparing spin rate, velocity, and break characteristics across clusters",
    theme = theme(
      plot.title = element_text(size = 18, face = "bold", hjust = 0.5),
      plot.subtitle = element_text(size = 12, hjust = 0.5, color = "gray30")
    )
  )
```



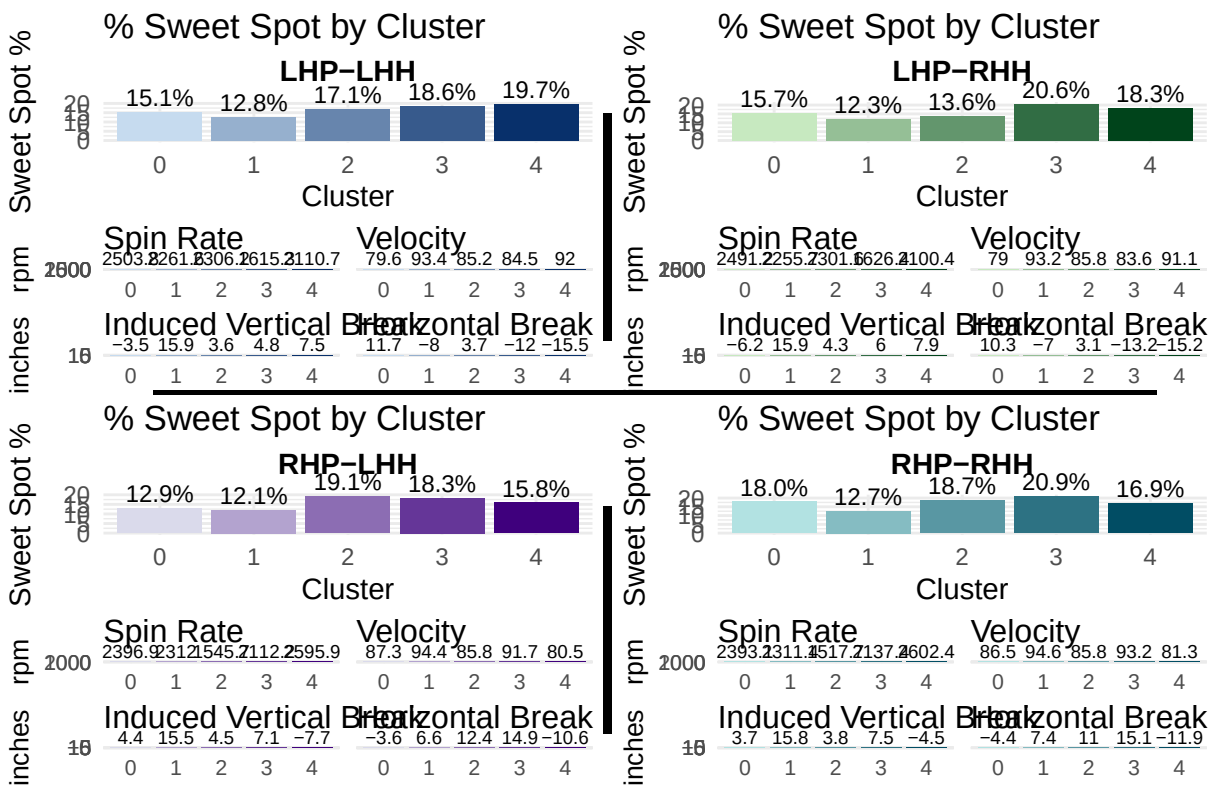

```
# --- Libraries ---
library(dplyr)
library(tidyr)
library(ggplot2)
library(patchwork)
library(scales)
```

```
##
## Attaching package: 'scales'

## The following object is masked from 'package:purrr':
##
##     discard

## The following object is masked from 'package:readr':
##
##     col_factor
```

```
# ---------- Define custom gradient palettes (light → dark) ----------
pal_LHPLHH <- colorRampPalette(c("#c6dbef", "#08306b"))(5)   # blue shades
pal_LHPRHH <- colorRampPalette(c("#c7e9c0", "#00441b"))(5)   # green shades
pal_RHPLHH <- colorRampPalette(c("#dadaeb", "#3f007d"))(5)   # purple shades
pal_RHPRHH <- colorRampPalette(c("#b2e2e2", "#014d64"))(5)   # teal shades
```

```r
# ---------- Sweet-spot plot helper (per matchup) ----------
sweet_plot <- function(df, label, pal){
  df %>%
    group_by(Cluster) %>%
    summarise(
      n_pitches = n(),
      n_sweetspot = sum(GIDP_SweetSpot == 1, na.rm = TRUE),
      pct_sweetspot = 100 * n_sweetspot / n_pitches,
      .groups = "drop"
    ) %>%
    mutate(Cluster = factor(Cluster)) %>%
    ggplot(aes(x = Cluster, y = pct_sweetspot, fill = Cluster)) +
    geom_col(show.legend = FALSE) +
    geom_text(aes(label = sprintf("%.1f%%", pct_sweetspot)),
              vjust = -0.35, size = 3) +
    scale_fill_manual(values = pal) +
    labs(title = "% Sweet Spot by Cluster", subtitle = label,
         x = "Cluster", y = "Sweet Spot %") +
    theme_minimal() +
    coord_cartesian(ylim = c(0, NA), clip = "off") +
    theme(
      plot.subtitle = element_text(face = "bold", hjust = 0.5),
      plot.margin = margin(4, 4, 4, 4)
    )
}

# ---------- 4-metric patchwork helper (per matchup) ----------
make_metric_patchwork <- function(df, pal){
  summary_df <- df %>%
    group_by(Cluster) %>%
    summarise(
      spinrate_mean         = mean(spinrate, na.rm = TRUE),
      relspeed_mean         = mean(relspeed, na.rm = TRUE),
      inducedvertbreak_mean = mean(inducedvertbreak, na.rm = TRUE),
      horzbreak_mean        = mean(horzbreak, na.rm = TRUE),
      .groups = "drop"
    ) %>%
    pivot_longer(-Cluster, names_to = "metric", values_to = "value") %>%
    mutate(Cluster = factor(Cluster))

  make_bar <- function(metric_name, title, ylab){
    ggplot(filter(summary_df, metric == metric_name),
           aes(x = Cluster, y = value, fill = Cluster)) +
      geom_col(show.legend = FALSE) +
      geom_text(aes(label = round(value, 1)), vjust = -0.3, size = 2.5) +
      scale_fill_manual(values = pal) +
      labs(title = title, x = NULL, y = ylab) +
      theme_minimal(base_size = 10) +
      coord_cartesian(clip = "off") +
      theme(plot.margin = margin(2,2,2,2))
  }

  p1 <- make_bar("spinrate_mean",         "Spin Rate",              "rpm")
```

```
  p2 <- make_bar("relspeed_mean",         "Velocity",                "mph")
  p3 <- make_bar("inducedvertbreak_mean", "Induced Vertical Break", "inches")
  p4 <- make_bar("horzbreak_mean",        "Horizontal Break",        "inches")

  p2 <- p2 + theme(axis.title.y = element_blank(),
                   axis.text.y  = element_blank(),
                   axis.ticks.y = element_blank())
  p4 <- p4 + theme(axis.title.y = element_blank(),
                   axis.text.y  = element_blank(),
                   axis.ticks.y = element_blank())

  (p1 | p2) / (p3 | p4)
}

# ---------- Build each matchup stack (sweet-spot + metrics) ----------
stack_LHPLHH <- sweet_plot(LHPLHH, "LHP-LHH", pal_LHPLHH) / make_metric_patchwork(LHPLHH, pal_LHPLHH) +
stack_LHPRHH <- sweet_plot(LHPRHH, "LHP-RHH", pal_LHPRHH) / make_metric_patchwork(LHPRHH, pal_LHPRHH) +
stack_RHPLHH <- sweet_plot(RHPLHH, "RHP-LHH", pal_RHPLHH) / make_metric_patchwork(RHPLHH, pal_RHPLHH) +
stack_RHPRHH <- sweet_plot(RHPRHH, "RHP-RHH", pal_RHPRHH) / make_metric_patchwork(RHPRHH, pal_RHPRHH) +

# ---------- "+" divider setup ----------
v_div <- ggplot() + geom_rect(aes(xmin=0, xmax=1, ymin=0, ymax=1), fill="black") + theme_void()
h_div <- ggplot() + geom_rect(aes(xmin=0, xmax=1, ymin=0, ymax=1), fill="black") + theme_void()

v_w <- 0.02
h_h <- 0.02

# ---------- Assemble full layout ----------
top_row    <- (stack_LHPLHH | v_div | stack_LHPRHH) + plot_layout(widths = c(1, v_w, 1))
bottom_row <- (stack_RHPLHH | v_div | stack_RHPRHH) + plot_layout(widths = c(1, v_w, 1))

final_plot <-
  (top_row / h_div / bottom_row) +
  plot_layout(heights = c(1, h_h, 1)) +
  plot_annotation(
    title = "GIDP Sweet Spots and Cluster Profiles by Matchup",
    theme = theme(plot.title = element_text(hjust = 0.5, face = "bold", size = 16))
  )

# ---------- Show & Save ----------
final_plot
```




```
ggsave("gidp_matchup_newcolors.png", final_plot, width = 14, height = 12, dpi = 300)
```

Including Plots

```
library(dplyr)
```