Permutation

Group E

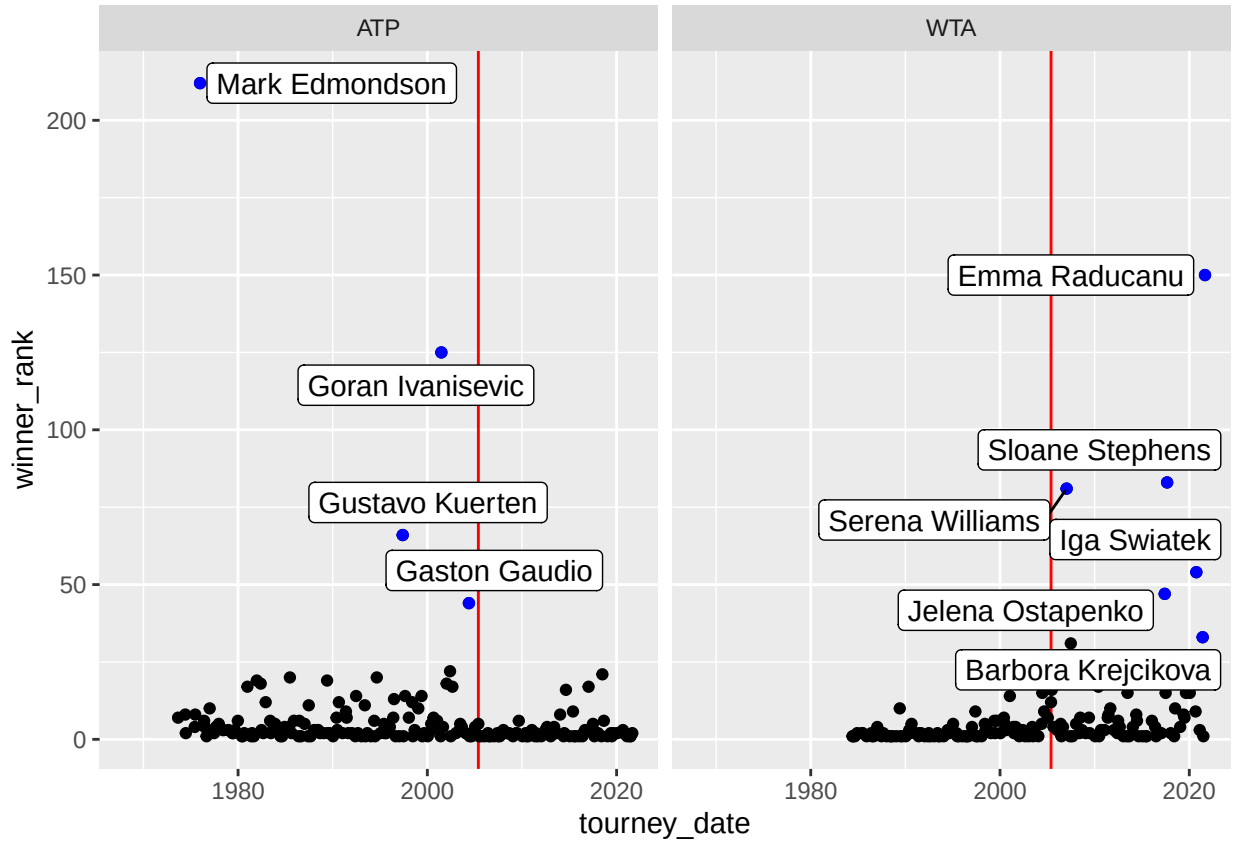
12/10/2021

```
library(ggrepel)
rank_point_results <- tennis_results %>%
  filter(round == "F",
         tourney_level == "G")

highlight_rank_point <- tennis_results %>%
  filter(round == "F",
         tourney_level == "G",
         winner_rank > 32)

rank_point_results %>%
  ggplot(aes(x = tourney_date, y = winner_rank)) +
  geom_vline(xintercept = rank_point_results$tourney_date[150], # split at 2005
             color = "red") +
  geom_point() +
  geom_point(data = highlight_rank_point,
             color = "blue") +
  facet_grid(~ tour) +
  geom_label_repel(data = highlight_rank_point,
                   aes(label = winner_name))
```

```
## Warning: Removed 95 rows containing missing values (geom_point).
```

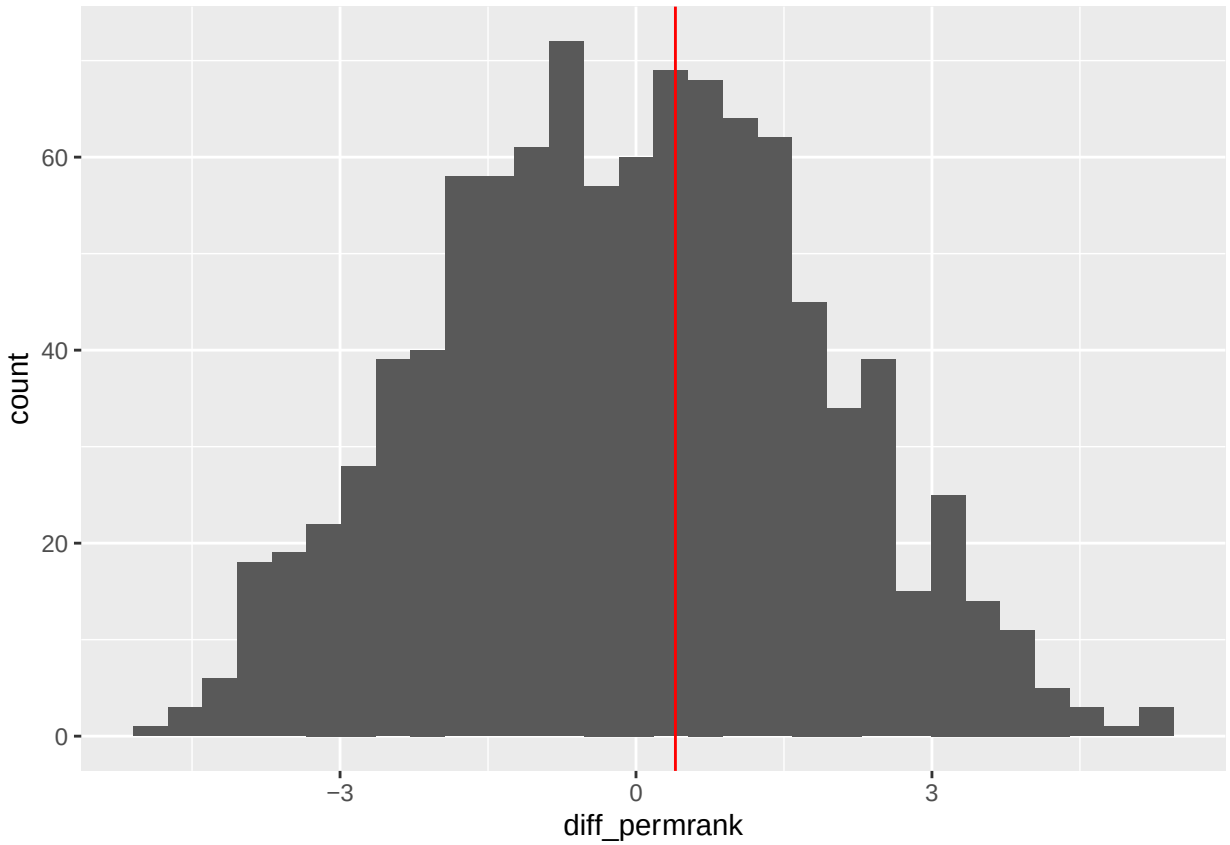



```
rank_point_results <- tennis_results %>%
  filter(round == "F",
         tourney_level == "G")

diff_rank_func <- function(.x){
  rank_point_results %>%
  filter(!is.na(winner_rank)) %>%
  mutate(permrank = sample(winner_rank, replace = FALSE)) %>%
  group_by(tour) %>%
  summarize(avg_permrank = mean(permrank),
            avg_rank = mean(winner_rank)) %>%
  summarize(diff_permrank = diff(avg_permrank),
            diff_rank = diff(avg_rank))
  }

set.seed(47)
perm_diff_rank <- map_df(1:1000, diff_rank_func)
perm_diff_rank %>%
  ggplot() +
  geom_histogram(aes(x = diff_permrank)) +
  geom_vline(aes(xintercept = diff_rank), color = "red")
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```




```
perm_diff_rank %>%
  summarize(pval = sum(diff_rank > diff_permrank) / 1000)
```

```
## # A tibble: 1 x 1
##    pval
##   <dbl>
## 1 0.589
```