Prediction

Youlan Shen

2023-04-29

Task 4: Prediction Using Estimated Model Parameters

```
library(tidyverse)
library(dplyr)
library(cowplot)
```

load data

```
load("hurricane.RData")
data <- hurricane %>% ungroup() %>%
  mutate(Season = Season - 1949)
load("test2.RData")
B <- test2$B
gamma <- test2$gamma
```

```
# get XYD
Y <- as.matrix(data$Wind.kt)
D <- data %>% mutate(intercept = 1) %>% dplyr::select(intercept, Wind_prev,
                                                      Lat_change, Long_change,
                                                      Wind_change) %>%
  as.matrix()
X <- data %>%
  mutate(index = 1:nrow(data)) %>%
  group_by(index) %>%
  dplyr::select(Month, Season, Nature) %>%
  mutate(n = 1) %>%
  pivot_wider(names_from = Month, values_from = n, values_fill = 0) %>%
  mutate(n = 1) %>%
  pivot_wider(names_from = Nature, values_from = n, values_fill = 0) %>%
  ungroup() %>%
  dplyr::select(April, May, June, July, August, September, October, November,
                December, Season, ET,NR,SS,TS) %>%
  as.matrix()
```

```
## Adding missing grouping variables: `index`
```

Get Estimated Model Parameters Using MCMC Train

```
# get parameter estimates function
get_estimate <- function(null_matrix, iter, MCMCchain){
  for (i in iter) {
    null_matrix <- null_matrix + MCMCchain[[i]]
  }
  estimate <- null_matrix / length(iter)
  estimate
}
# set iter, we only use the last 3000 of MCMC to get posterior mean
iter <- 3001:6000
# set Null matrix
beta_estimate <- matrix(data = 0, nrow = 5, ncol = 699)
gamma_estimate <- matrix(data = 0, nrow = 14, ncol = 1)
beta_estimate <- get_estimate(beta_estimate, iter, B)
gamma_estimate <- get_estimate(gamma_estimate, iter, gamma)
```

Get Prediction

```
X_with_id <- cbind(as.numeric(factor(data$ID)), X)
#number of hurricanes
H <- max(X_with_id[,1])
ith_hurricane_idx <- 1:H |>
  map(\(i) which(X_with_id[,1] == i))
# calculate Y estimates
N <- length(Y)
mu_H <- rep(NA, N)
for (i in 1:H) {
    curr_hurricane_idx <- ith_hurricane_idx[[i]]
    mu_i <- D[curr_hurricane_idx,,drop = FALSE] %*%
      beta_estimate[,i,drop = FALSE]
    mu_H[curr_hurricane_idx] <- mu_i
}
mu_H <- as.matrix(mu_H)
Wind_pred <- mu_H + X %*% gamma_estimate
```

```r
# get the posterior mean
colnames(beta_estimate) <- unique(data$ID)
colnames(gamma_estimate) <- "gamma_estimate"
beta_estimate[, 1:5] %>%
  knitr::kable(
    caption = "Random Effects Beta Estimates for Each Hurricane",
    digits = 3)
```

Table 1: Random Effects Beta Estimates for Each Hurricane

	ABLE.1950	BAKER.1950	CHARLIE.1950	DOG.1950	EASY.1950
intercept	4.751	4.704	4.497	4.372	4.516
Wind_prev	0.895	0.849	0.894	0.941	0.936
Lat_change	0.444	0.320	0.160	-0.167	0.078
Long_change	-0.490	-0.300	-0.352	-0.511	-0.384
Wind_change	0.456	0.354	0.470	0.487	0.546

```r
gamma_estimate %>%
  knitr::kable(
    caption = "Fixed Effects Gamma Estimates for Each Covariate",
    digits = 3
  )
```

Table 2: Fixed Effects Gamma Estimates for Each Covariate

	gamma_estimate
April	0.001
May	-0.001
June	0.002
July	-0.006
August	0.005
September	-0.026
October	-0.003
November	0.010
December	-0.004
Season	-0.021
ET	-0.129
NR	-0.007
SS	0.010
TS	0.196

Analysis on Prediction

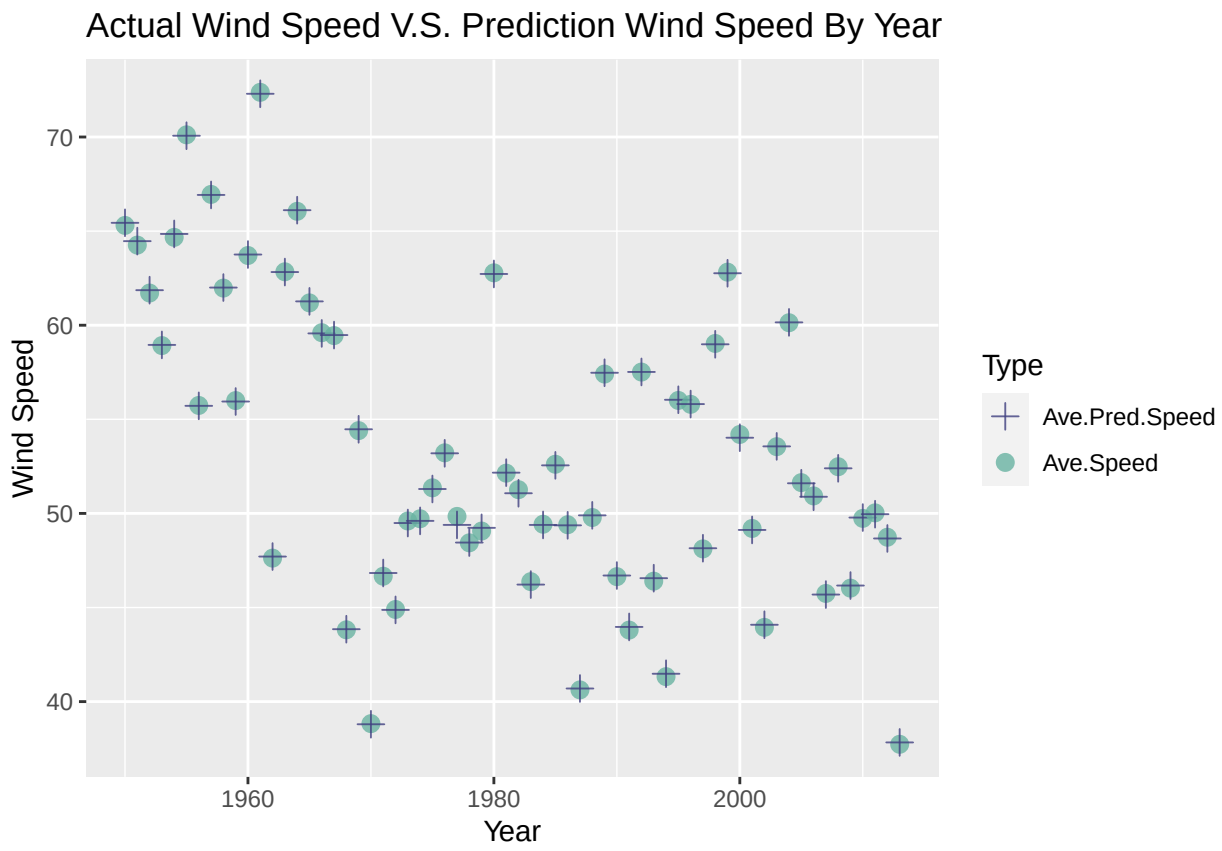
```r
hurricane["Wind_pred"] <- Wind_pred
hurricane <- hurricane %>%
  dplyr::select(ID, Season, Month, Nature, Date, Wind.kt, Wind_pred)
hurricane %>%
  mutate(Pred_E = Wind.kt - Wind_pred,
```

```
        Pred_E_Sq = Pred_E * Pred_E) %>%
  group_by(ID) %>%
  summarise(nobs = n(),
            Pred_RMSE = sqrt(sum(Pred_E_Sq)/nobs)) %>%
  arrange(Pred_RMSE) %>%
  head(10) %>%
  knitr::kable(
      caption = "Prediction Standard Error of Wind Speed in Ascending Order",
      col.names = c("Hurricane ID", "Number of Observation", "Prediction RMSE"),
      digits = 3
  )
```

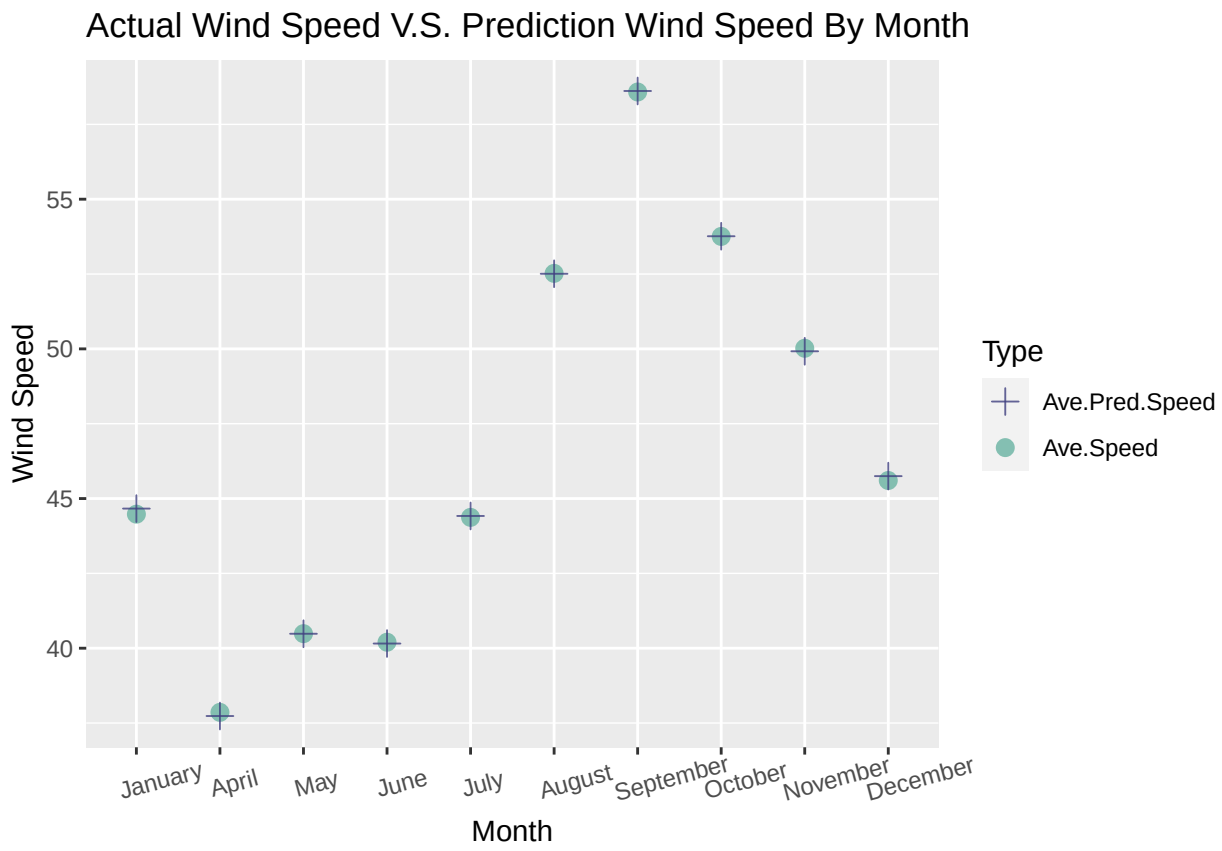
Table 3: Prediction Standard Error of Wind Speed in Ascending Order

Hurricane ID	Number of Observation	Prediction RMSE
SIXTEEN.2008	6	1.041
FABIAN.1991	9	1.259
FOUR:UNNAMED.1988	5	1.373
TEN.2005	19	1.422
LORENZO.2001	14	1.475
FIFTEEN:UNNAMED.1988	4	1.845
FABIAN.1997	15	1.939
ANDREW.1986	14	1.943
TEN.2011	7	1.984
CINDY.1987	20	1.997

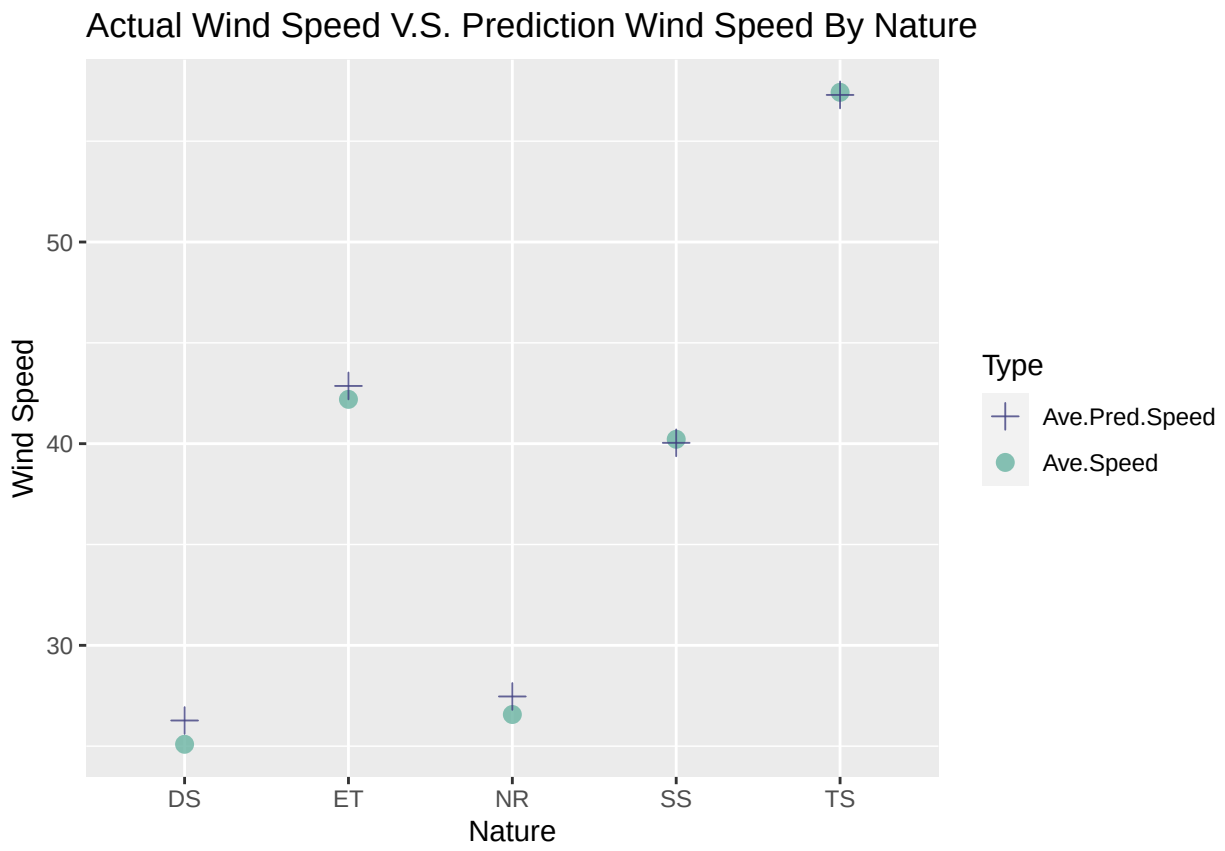
```
hurricane %>%
  group_by(Season) %>%
  summarise(Ave.Speed = mean(Wind.kt),
            Ave.Pred.Speed = mean(Wind_pred)) %>%
  pivot_longer(Ave.Speed:Ave.Pred.Speed,
               names_to = "Type",
               values_to = "Ave.Wind.Speed") %>%
  ggplot(aes(x = Season, y = Ave.Wind.Speed)) +
  geom_point(aes(shape = Type, color = Type), alpha = 0.8, size = 3) +
  scale_color_manual(values = c("#404080", "#69b3a2")) +
  scale_shape_manual(values = c(3, 16)) +
  labs(title = "Actual Wind Speed V.S. Prediction Wind Speed By Year",
       x = "Year",
       y = "Wind Speed")
```



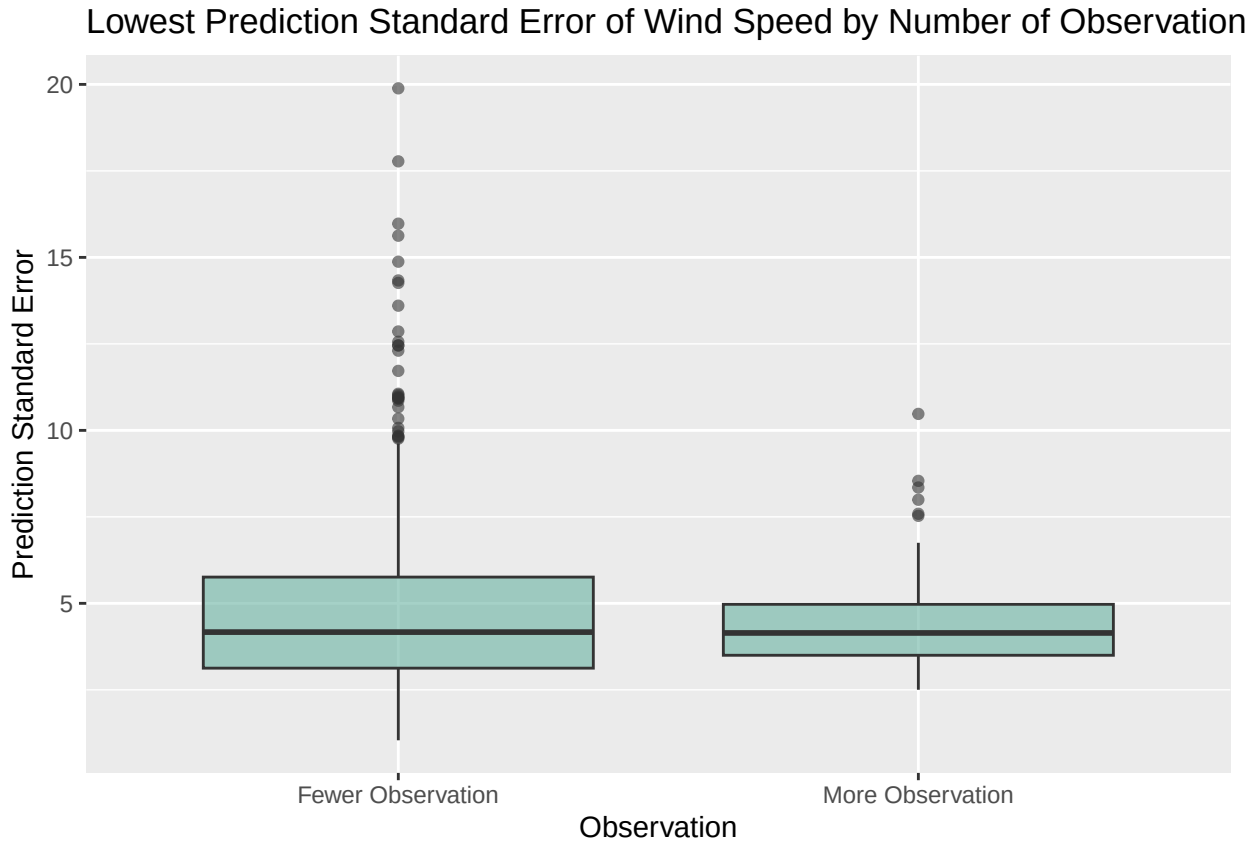

```
hurricane %>%
  group_by(Month) %>%
  summarise(Ave.Speed = mean(Wind.kt),
            Ave.Pred.Speed = mean(Wind_pred)) %>%
  pivot_longer(Ave.Speed:Ave.Pred.Speed,
               names_to = "Type",
               values_to = "Ave.Wind.Speed") %>%
  ggplot(aes(x = Month, y = Ave.Wind.Speed)) +
  geom_point(aes(shape = Type, color = Type), alpha = 0.8, size = 3) +
  scale_color_manual(values = c("#404080", "#69b3a2")) +
  scale_shape_manual(values = c(3, 16)) +
  labs(title = "Actual Wind Speed V.S. Prediction Wind Speed By Month",
       x = "Month",
       y = "Wind Speed") +
  theme(axis.text.x = element_text(angle = 15, vjust = 0.5, hjust = 0.2))
```



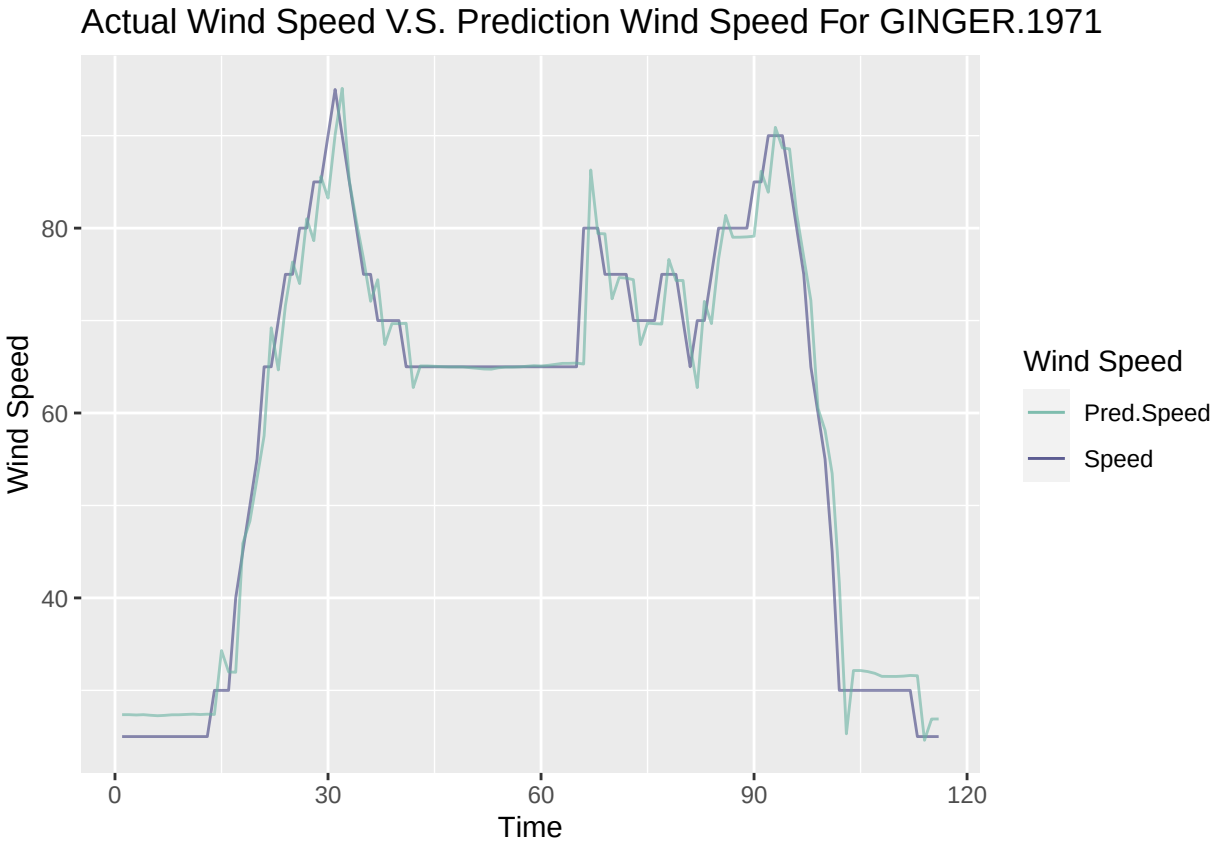

```
hurricane %>%
  group_by(Nature) %>%
  summarise(Ave.Speed = mean(Wind.kt),
            Ave.Pred.Speed = mean(Wind_pred)) %>%
  pivot_longer(Ave.Speed:Ave.Pred.Speed,
               names_to = "Type",
               values_to = "Ave.Wind.Speed") %>%
  ggplot(aes(x = Nature, y = Ave.Wind.Speed)) +
  geom_point(aes(shape = Type, color = Type), alpha = 0.8, size = 3) +
  scale_color_manual(values = c("#404080", "#69b3a2")) +
  scale_shape_manual(values = c(3, 16)) +
  labs(title = "Actual Wind Speed V.S. Prediction Wind Speed By Nature",
       x = "Nature",
       y = "Wind Speed")
```




```
hurricane %>%
  mutate(Pred_E = Wind.kt - Wind_pred,
         Pred_E_Sq = Pred_E * Pred_E) %>%
  group_by(ID) %>%
  summarise(n = n(),
            Pred_SE = sqrt(sum(Pred_E_Sq)/n)) %>%
  mutate(nobs = ifelse(n > 50, "More Observation", "Fewer Observation")) %>%
  group_by(nobs) %>%
  ggplot(aes(x = nobs, y = Pred_SE)) +
  geom_boxplot(fill = "#69b3a2", alpha = 0.6) +
  labs(title = "Lowest Prediction Standard Error of Wind Speed by Number of Observation",
       x = "Observation",
       y = "Prediction Standard Error")
```




```
hurricane %>%
  filter(ID == "GINGER.1971") %>%
  mutate(Obs = 1:116) %>%
  ggplot(aes(x = Obs)) +
  geom_line(aes(y = Wind.kt, color = "Speed"), alpha = 0.6) +
  geom_line(aes(y = Wind_pred, color = "Pred.Speed"), alpha = 0.6) +
  scale_color_manual(name = "Wind Speed",
                     values = c("Speed" = "#404080",
                                "Pred.Speed" = "#69b3a2")) +
  labs(title = "Actual Wind Speed V.S. Prediction Wind Speed For GINGER.1971",
       x = "Time",
       y = "Wind Speed")
```


Actual Wind Speed V.S. Prediction Wind Speed For GINGER.1971
Wind Speed
80
60
40
0
30
60
90
120
Time
Wind Speed
Pred.Speed
Speed

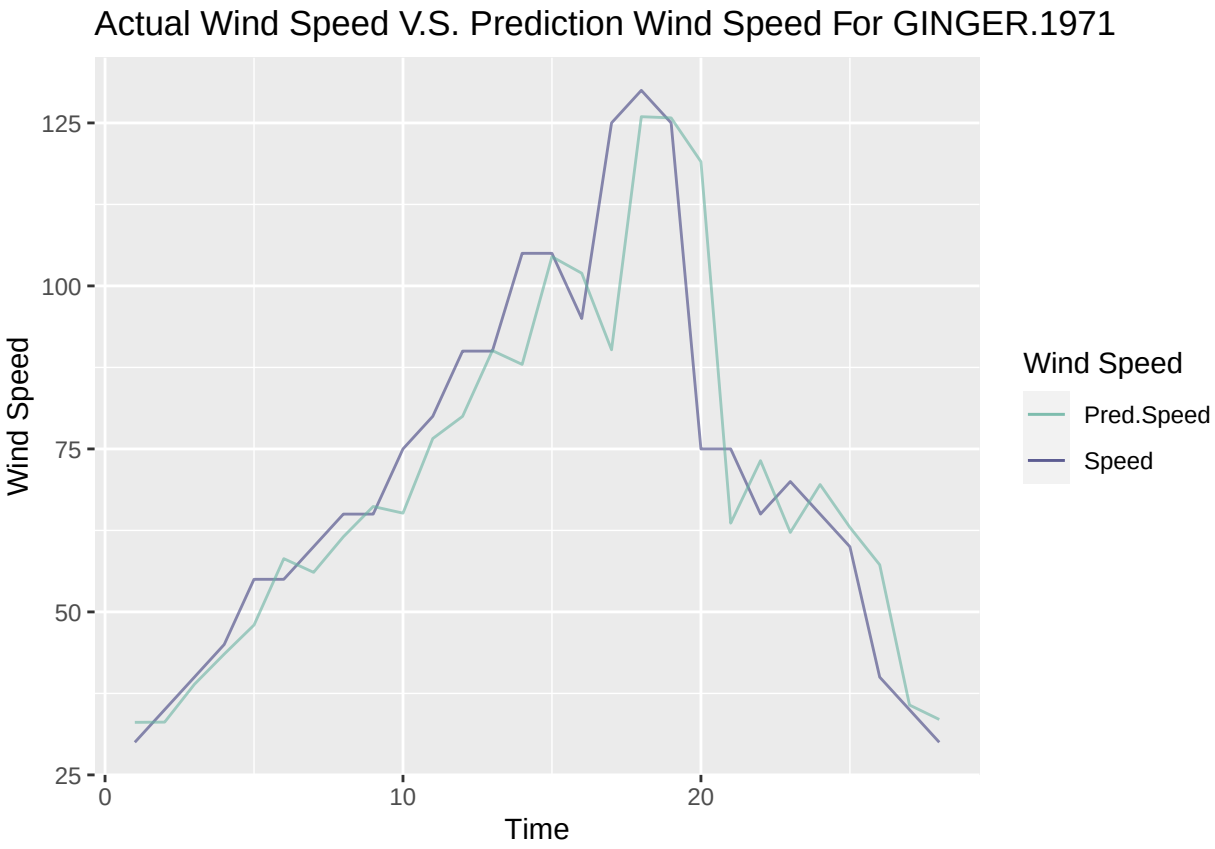
```r
hurricane %>%
  mutate(Pred_E = Wind.kt - Wind_pred,
         Pred_E_Sq = Pred_E * Pred_E) %>%
  group_by(ID) %>%
  summarise(nobs = n(),
            Pred_RMSE = sqrt(sum(Pred_E_Sq)/nobs)) %>%
  arrange(Pred_RMSE) %>%
  top_n(10) %>%
  knitr::kable(
      caption = "Highest Prediction Standard Error of Wind Speed in Ascending Order",
      col.names = c("Hurricane ID", "Number of Observation", "Prediction RMSE"),
      digits = 3
  )
```

```
## Selecting by Pred_RMSE
```

Table 4: Highest Prediction Standard Error of Wind Speed in Ascending Order

Hurricane ID	Number of Observation	Prediction RMSE
CHARLEY.2004	28	12.561
JANET.1955	33	12.858
BLANCHE.1969	8	13.604
EDOUARD.1984	5	14.262
CELIA.1970	22	14.334
FELIX.2007	26	14.875
HATTIE.1961	18	15.626
IRIS.2001	20	15.974
AUDREY.1957	15	17.776
ETHEL.1960	12	19.886

```r
hurricane %>%
  filter(ID == "CHARLEY.2004") %>%
  mutate(Obs = 1:28) %>%
  ggplot(aes(x = Obs)) +
  geom_line(aes(y = Wind.kt, color = "Speed"), alpha = 0.6) +
  geom_line(aes(y = Wind_pred, color = "Pred.Speed"), alpha = 0.6) +
  scale_color_manual(name = "Wind Speed",
                     values = c("Speed" = "#404080",
                                "Pred.Speed" = "#69b3a2")) +
  labs(title = "Actual Wind Speed V.S. Prediction Wind Speed For GINGER.1971",
       x = "Time",
       y = "Wind Speed")
```


Actual Wind Speed V.S. Prediction Wind Speed For GINGER.1971
Wind Speed
125
100
75
50
25
0
10
20
Time
Wind Speed
Pred.Speed
Speed