

Assignment 1

Statistical Modelling: Theory and Practice

Laura Sans, Felix Pacheco, Begoña Bolós

10/4/2020

Project 1

WIND POWER FORECAST

1. Descriptive statistics

Set working directory and read the space-separated file

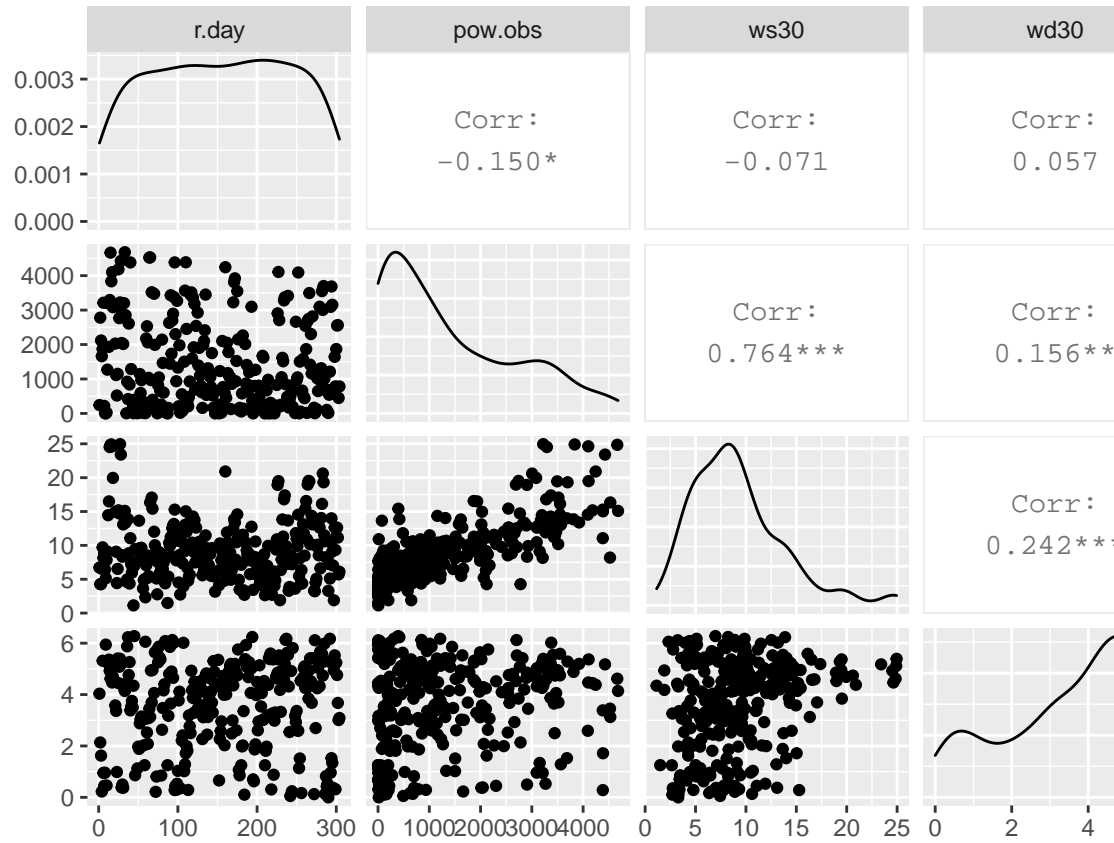
```
setwd(wd)
raw_wp <- read.csv("project_data/tuno.txt", sep=" ")
```

Summary statistics Below, the summary statistics for the Wind power production (pow.obs), the wind speed (ws30) and wind direction (wd30). The other three variables are categorical: r-days corresponds to the number of days from the start of the measurement. Month and day correspond to the date of the measurement. The start date is the 1st of January 2003 and the last day is 31st of October 2003, a total of 304 days.

```
summary(raw_wp[c("pow.obs", "ws30", "wd30")])
```

##	pow.obs	ws30	wd30
## Min.	: 0.123	Min. : 1.139	Min. :0.000095
## 1st Qu.	: 254.158	1st Qu.: 5.779	1st Qu.:2.474999
## Median	: 964.123	Median : 8.498	Median :4.079297
## Mean	:1381.196	Mean : 9.112	Mean :3.602390
## 3rd Qu.	:2196.579	3rd Qu.:11.202	3rd Qu.:4.945443
## Max.	:4681.062	Max. :24.950	Max. :6.274642

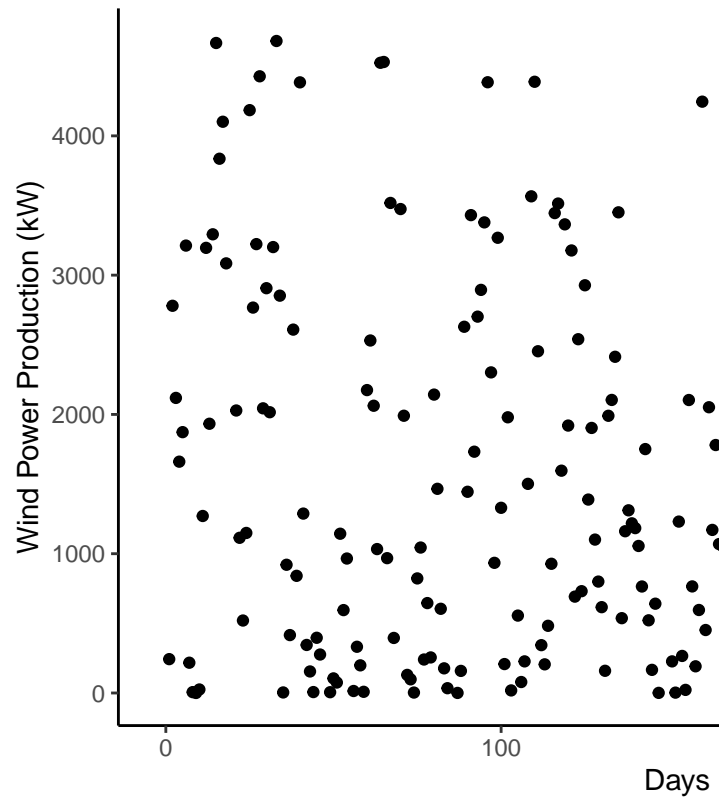
```
wp <- raw_wp %>% mutate_at(vars(month, day), factor)
ggpairs(data = wp[c("r.day", "pow.obs", "ws30", "wd30")],)
```



Pairs plot of all the data

```
ggplot(data=wp, aes(x=r.day, y=pow.obs)) + geom_point() + labs(title= "Distribution of wind power prod
```

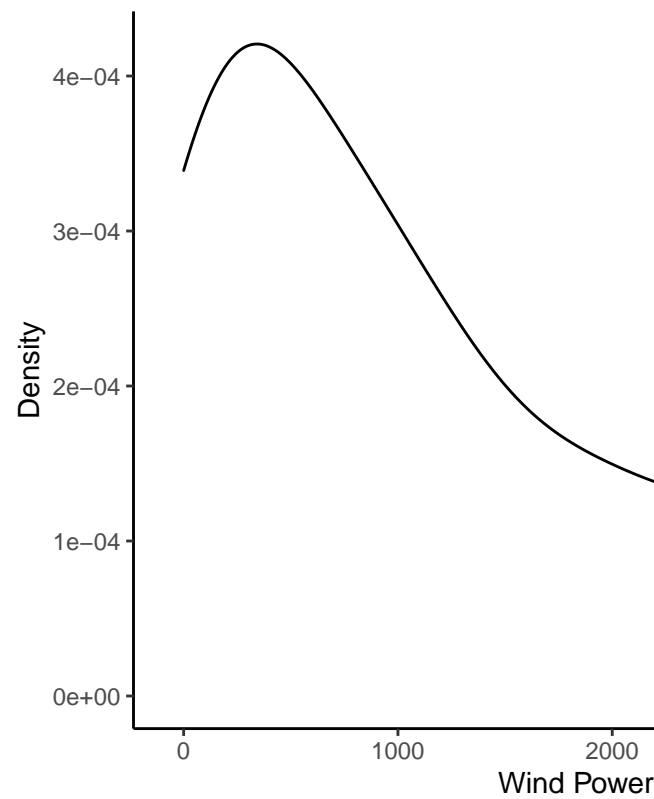
Distribution of wind power production along the time



Distribution of wind power production along the time

```
ggplot(data = wp, aes(x=pow.obs)) + geom_density() + labs(title= "Probability density plot of wind power")
```

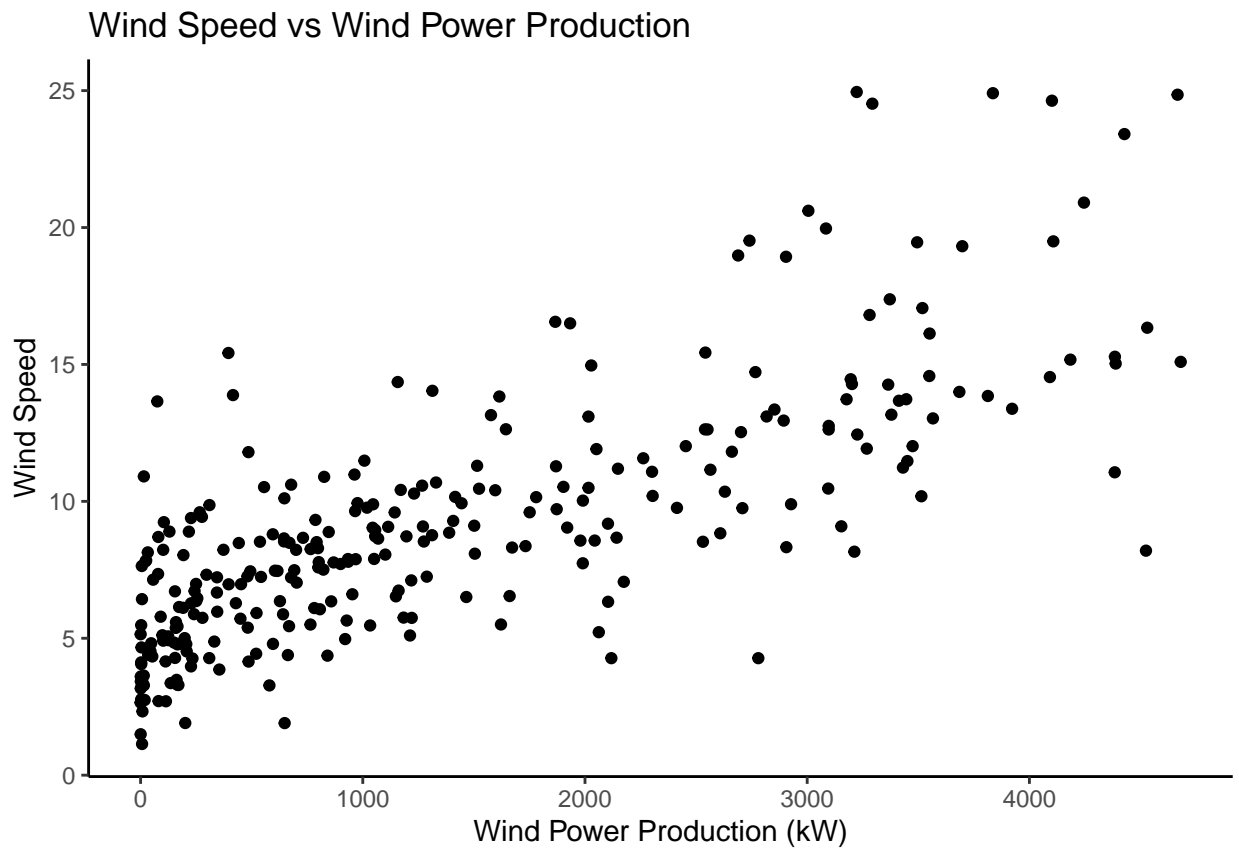
Probability density plot of wind power



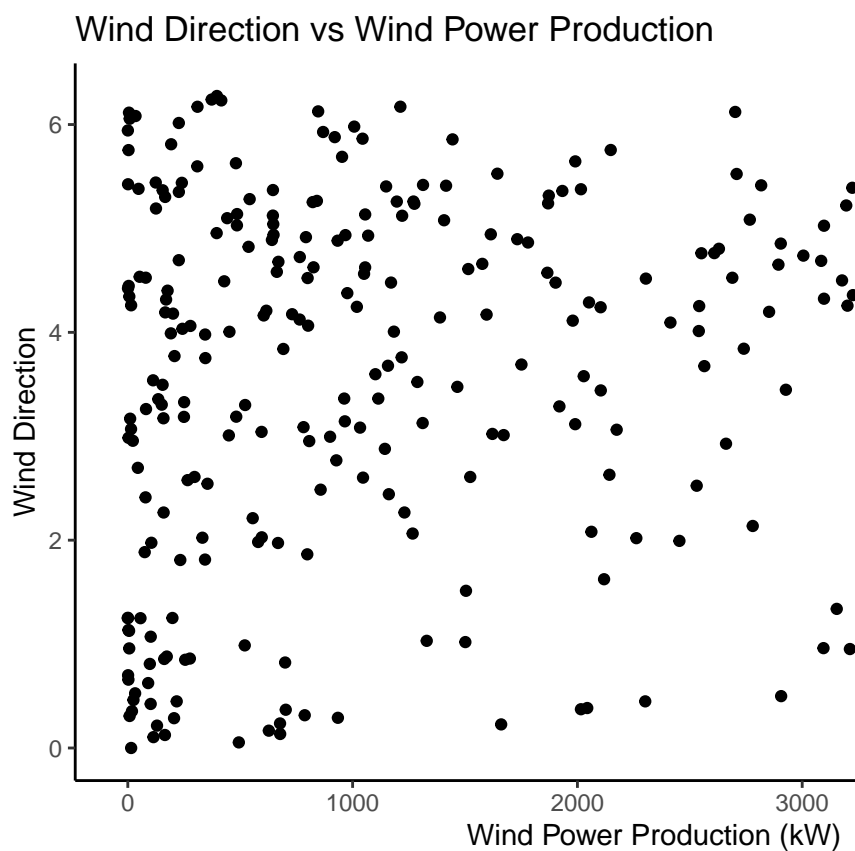
Probability density function of the Wind Power Production

Wind Speed vs Wind Power Production

```
ggplot(data = wp, aes(x=pow.obs, y=ws30)) + geom_point() + labs(title= "Wind Speed vs Wind Power Production")
```



```
# NOT SUPER INFORMATIVE  
ggplot(data = wp, aes(x=pow.obs, y=wd30)) + geom_point() + labs(title= "Wind Direction vs Wind Power Production")
```



Wind direction vs Wind Power Production

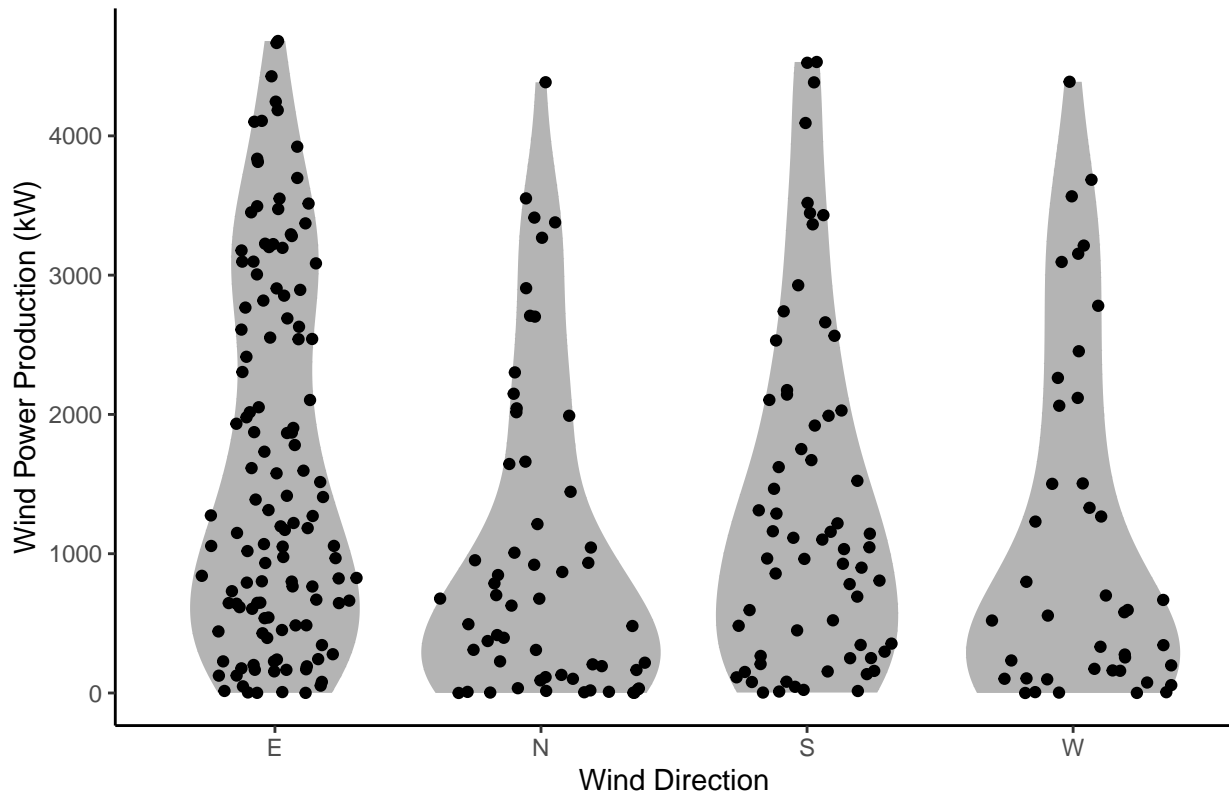
use case when

```
wp <- wp %>%
```

```
  mutate(direction = case_when( pi/4 >= wd30 ~ "N",
                                (7*pi)/4 < wd30 ~ "N",
                                (3*pi)/4 >= wd30 & wd30 > pi/4 ~ "W",
                                (5*pi)/4 >= wd30 & wd30 > (3*pi)/4 ~ "S",
                                (7*pi)/4 >= wd30 & wd30 > (5*pi)/4 ~ "E"))
```

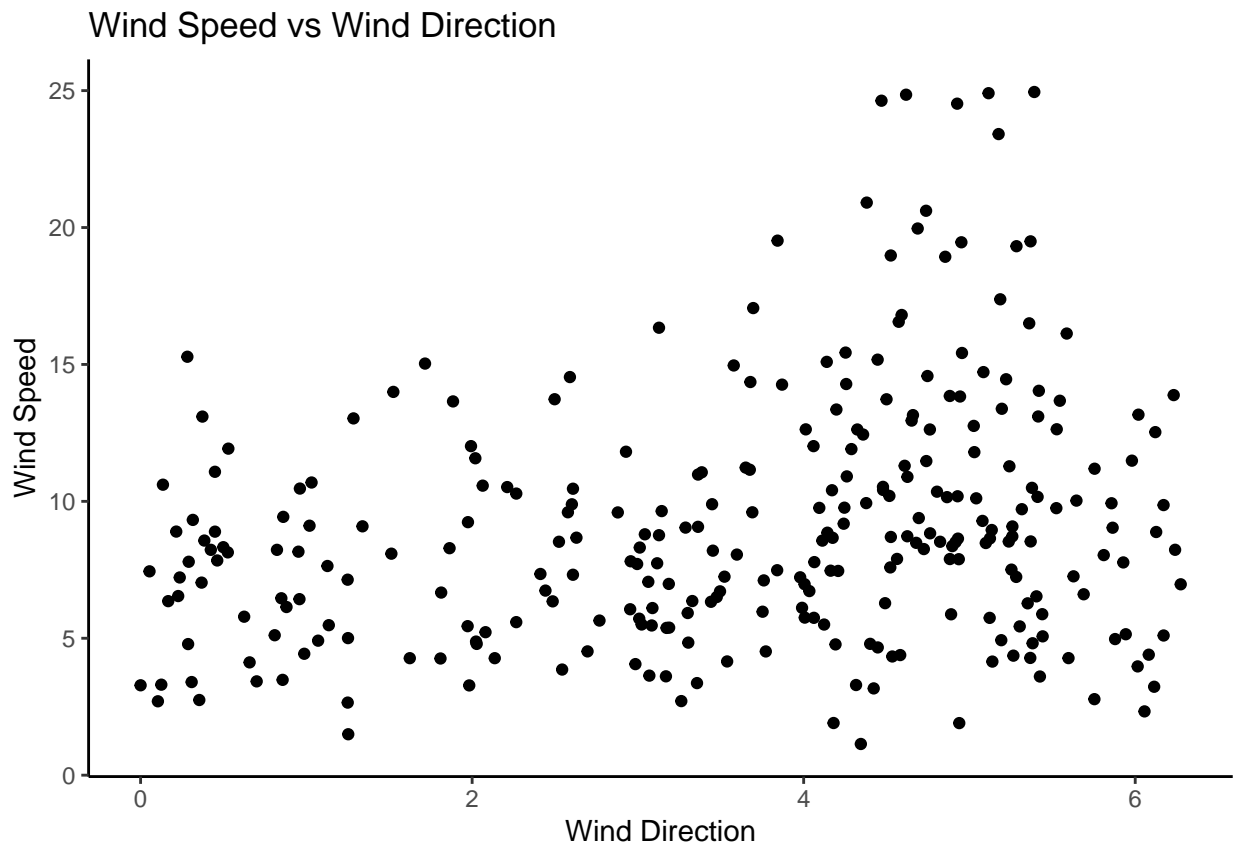
```
ggplot(data=wp, aes(x=direction, y=pow.obs)) + geom_violin(color=NA, fill="black", alpha= 0.3, draw_quantiles=c(0.25, 0.5, 0.75))
```

Wind direction effects on Wind Power Production

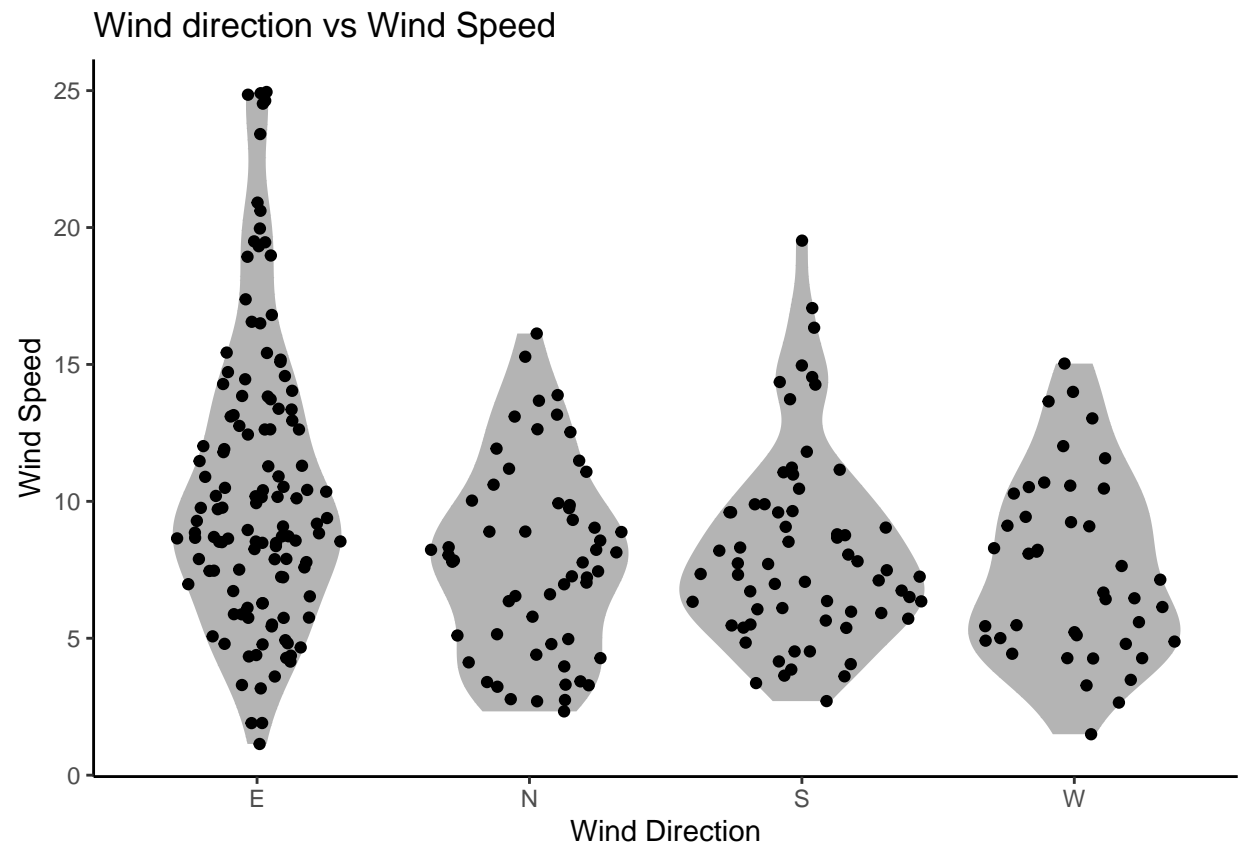


Other plots with wind direction

```
ggplot(data = wp, aes(y=ws30, x=wd30)) + geom_point() + labs(title= "Wind Speed vs Wind Direction", x =
```

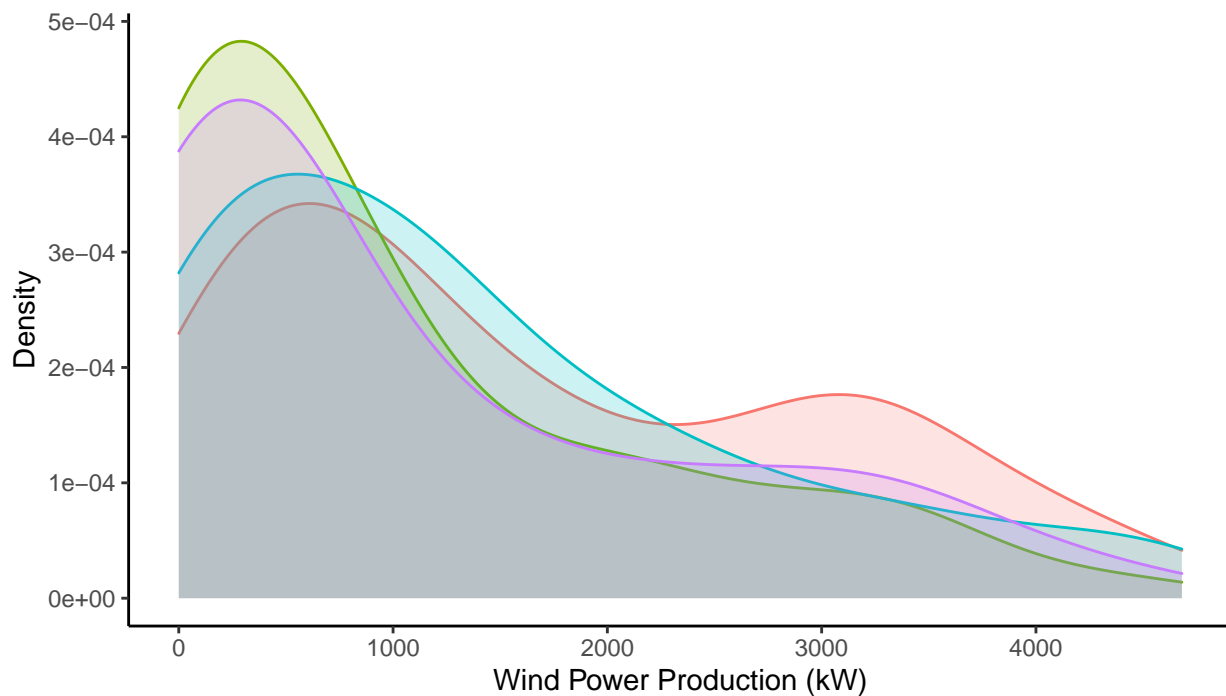


```
ggplot(data=wp, aes(x=direction, y=ws30)) + geom_violin(color=NA, fill="black", alpha= 0.3, draw_quantiles=)
```

```
ggplot(data = wp, aes(x=pow.obs, fill=direction, color = direction)) + geom_density(alpha =0.2) + labs
```

Probability density plot of wind power production



direction ■ E ■ N ■ S ■ W

```
# summary by direction
```

```
wp %>% group_by(direction) %>% count()
```

```
## # A tibble: 4 x 2
## # Groups:   direction [4]
##   direction     n
##   <chr>      <int>
## 1 E          124
## 2 N           57
## 3 S           65
## 4 W           42
```

```
wp %>% group_by(direction) %>% summarise(max = max(pow.obs))
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```
## # A tibble: 4 x 2
##   direction  max
##   <chr>      <dbl>
## 1 E        4681.
## 2 N        4385.
## 3 S        4530.
## 4 W        4388.
```

2. Simple models

Project 2

SURVIVAL DATA

Project 3

FINANCIAL DATA

1. DESCRIPTIVE STATISTICS AND SIMPLE MODELS