R & RStudio

Tools for Contemporary Analyses & Data Communication

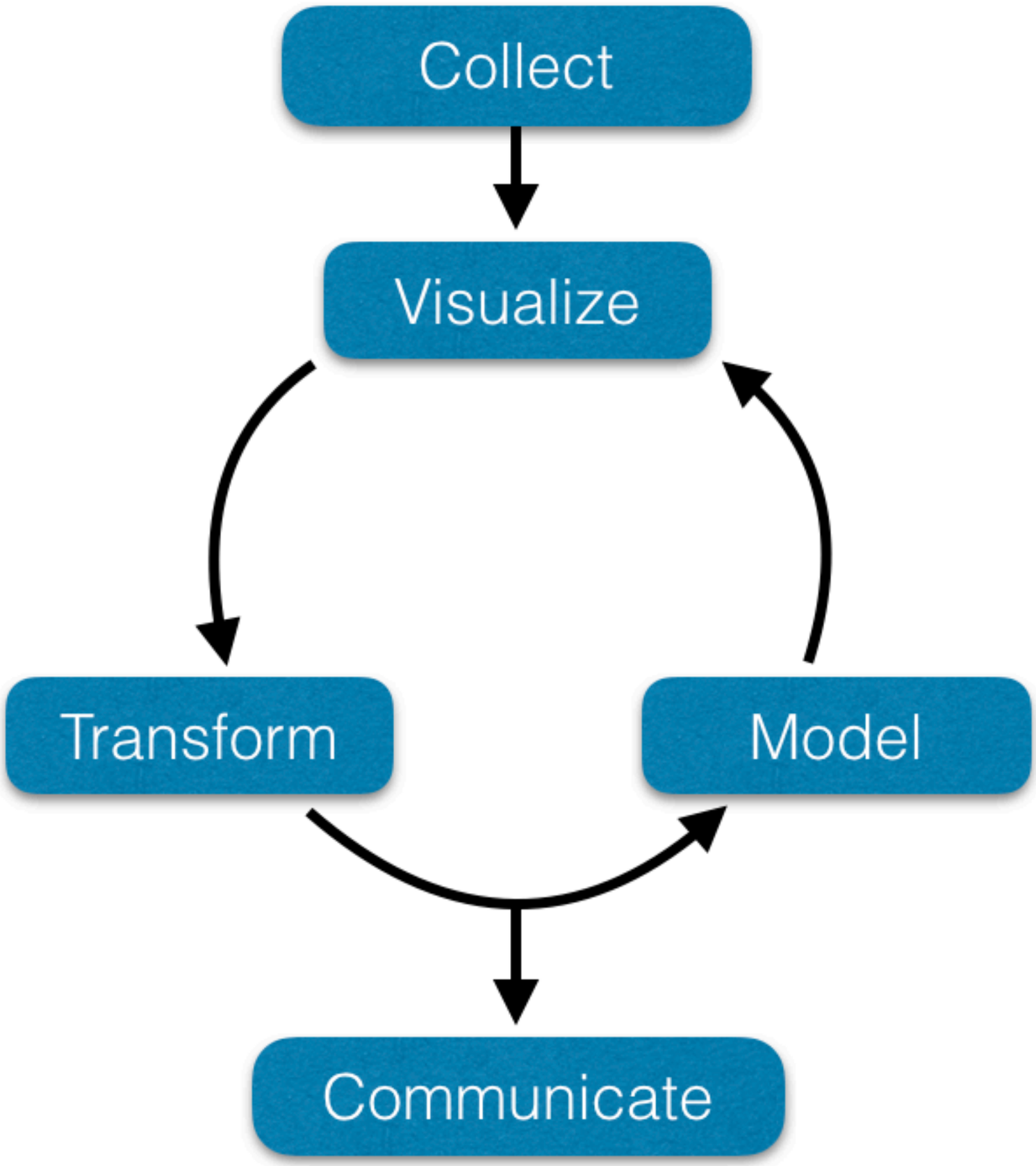
R. Dyer

https://rodneydyer.com

Why are you taking this class?

“Exploratory Data Analysis” is an attitude, a state of flexibility, a willingness to look for those things that we believe are not there, as well as those we believe to be there.

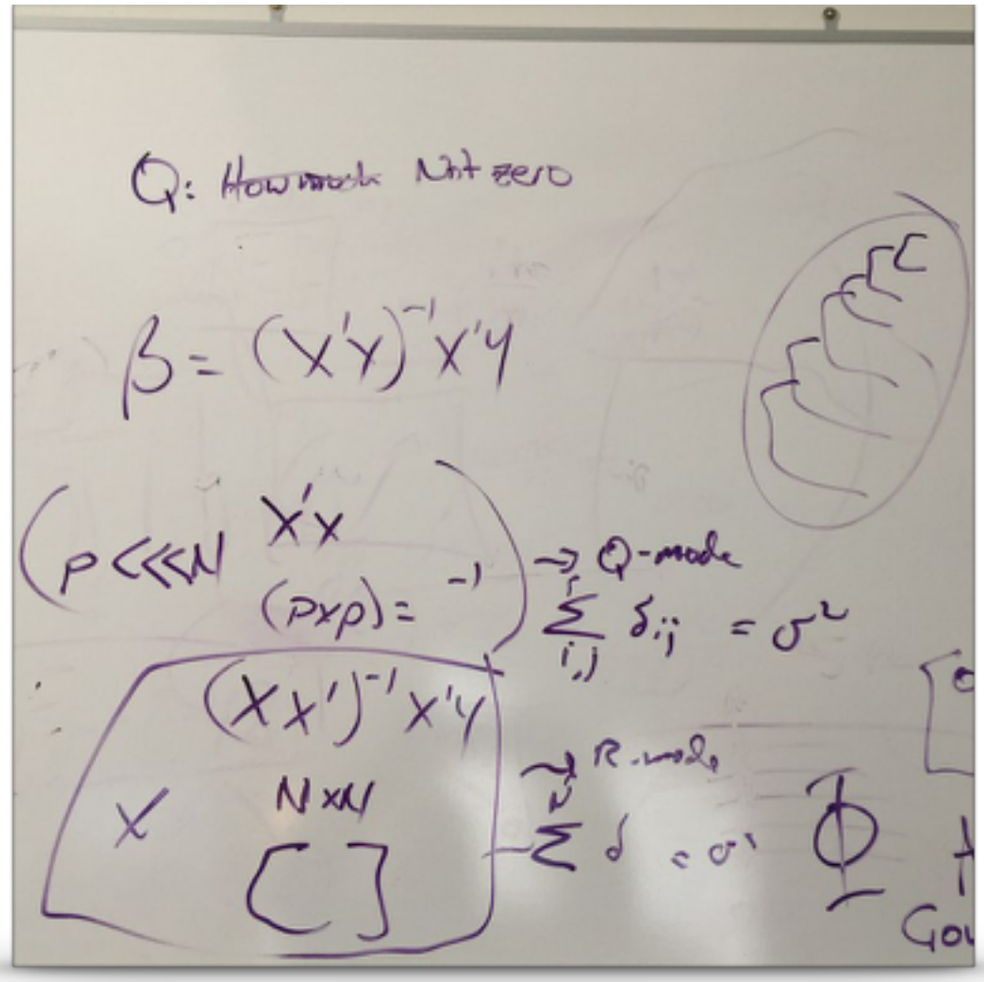
— John Tukey


Collect
Visualize
Transform
Model
Communicate

Quantitative Tools

Analytical Platforms - Pens

- Mastery required *a priori*
- Limitless in extent (*c.f.*, theoretical graduate theses)
- Rather steep learning curve…

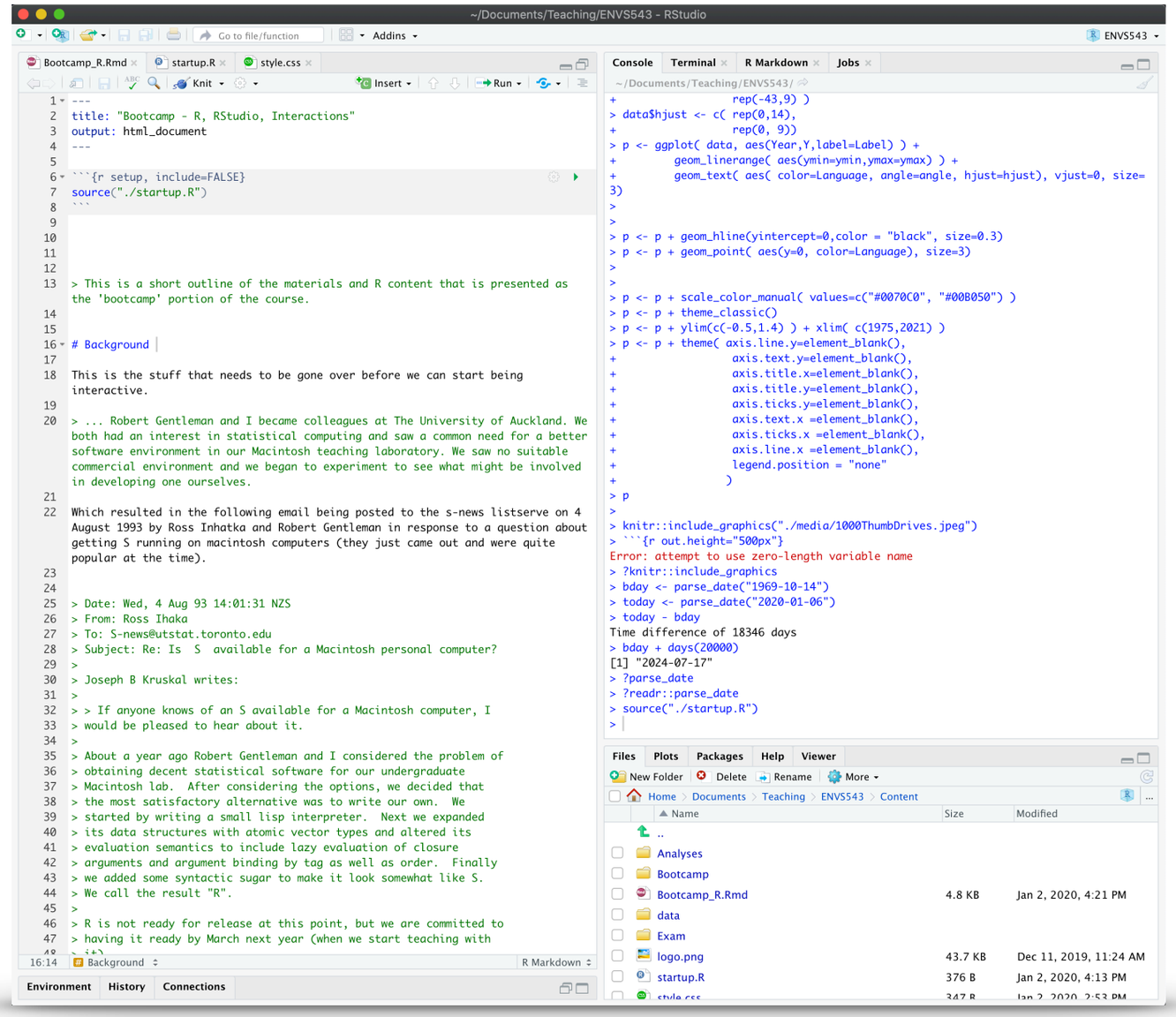


Point & Clicky

- Ease of use
- Proprietary
- Only common implementations
- Not scalable
- One-trick pony

Analysis Platforms

- Generalized tools
- Expandable
- Lower barrier to entry
- Broad applicability
- Collaborative
- Reproducible



Platform Grammars

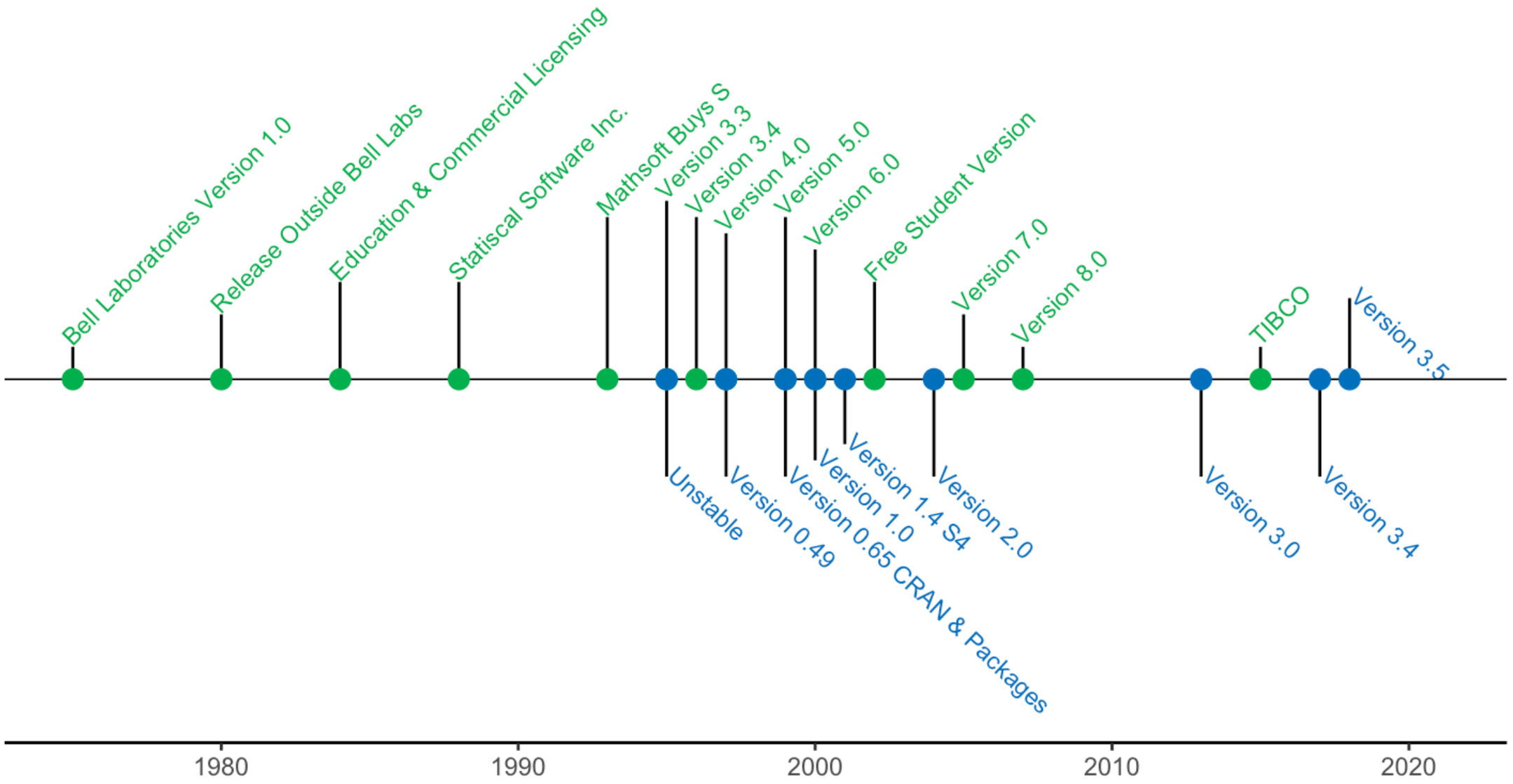
Which Tool Should I Learn?

Difference between R and Python

Parameter	R	Python
Objective	Data analysis and statistics	Deployment and production
Primary Users	Scholar and R&D	Programmers and developers
Flexibility	Easy to use available library	Easy to construct new models from scratch. I.e., matrix computation and optimization
Learning curve	Difficult at the beginning	Linear and smooth
Popularity of Programming Language. Percentage change	4.23% in 2018	21.69% in 2018
Average Salary	$99.000	$100.000
Integration	Run locally	Well-integrated with app
Task	Easy to get primary results	Good to deploy algorithm
Database size	Handle huge size	Handle huge size
IDE	Rstudio	Spyder, Ipthon Notebook
Important Packages and library	tydiverse, ggplot2, caret, zoo	pandas, scipy, scikit-learn, TensorFlow, caret
Disadvantages	Slow High Learning curve Dependencies between library	Not as many libraries as R
Advantages	• Graphs are made to talk. R makes it beautiful • Large catalog for data analysis • GitHub interface • RMarkdown • Shiny	• Jupyter notebook: Notebooks help to share data with colleagues • Mathematical computation • Deployment • Code Readability • Speed • Function in Python

https://www.guru99.com/r-vs-python.html

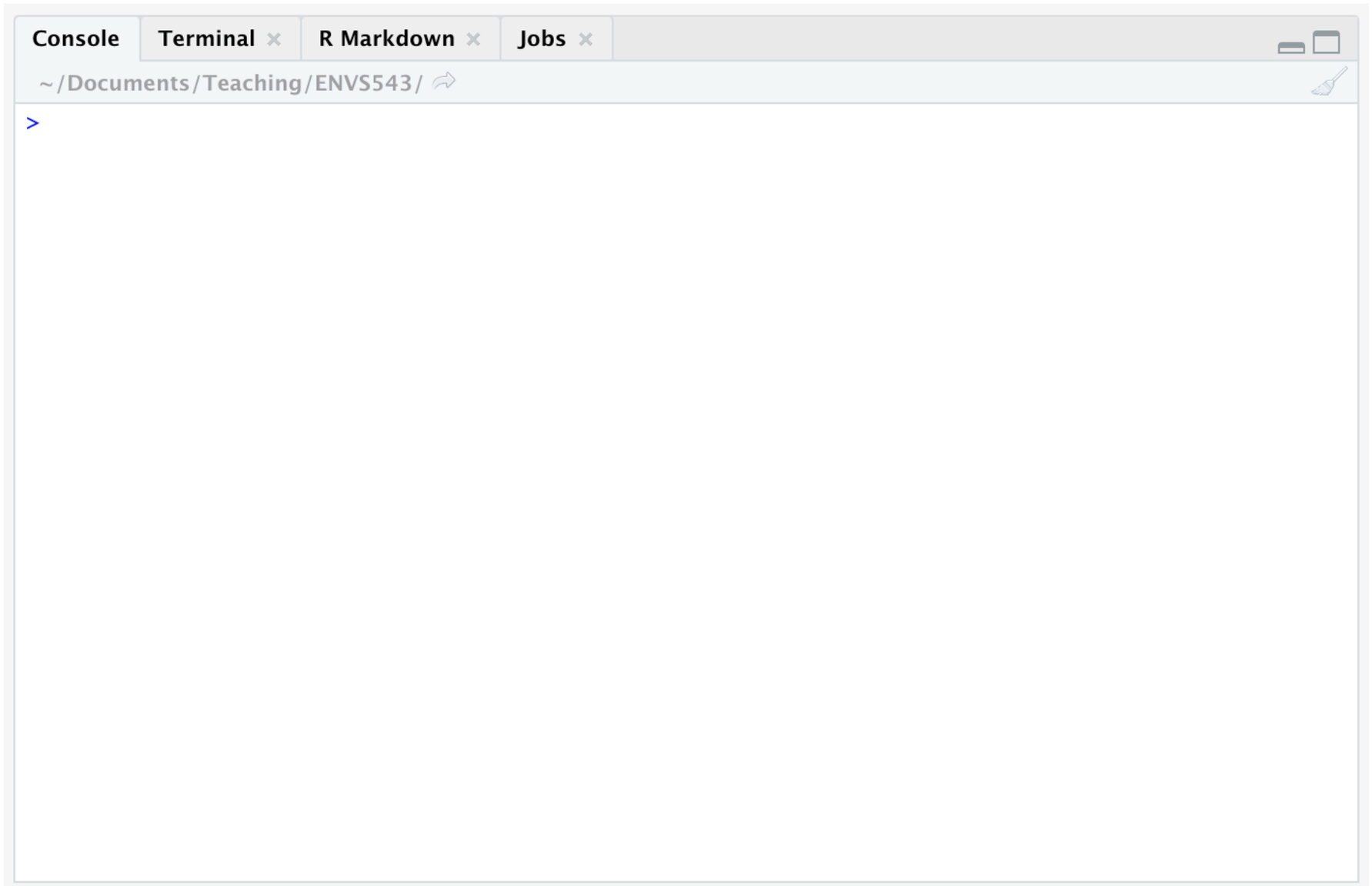
The Derivation of R



R Today

- A stable grammar
- An open source interpreter
- A public interface for expansion
- Available to everyone
- Most common platform for environmental sciences
- Interact through RStudio

Interactive Terminal



Very effective way to do ‘quick’ calculations.

Type & Look Interaction

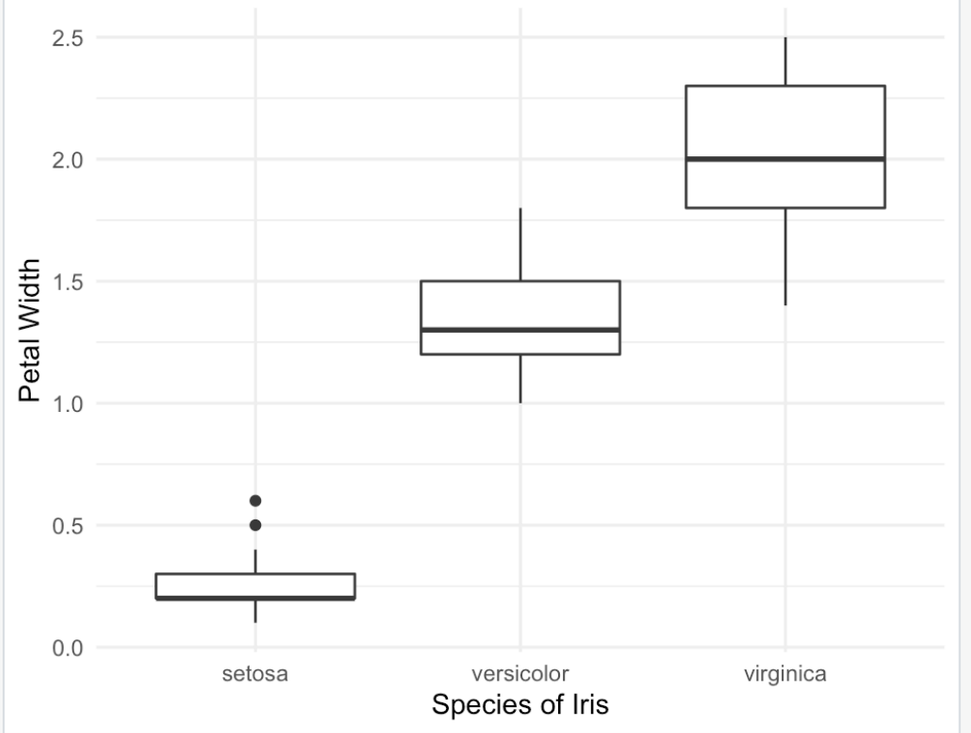
Console | Terminal | Jobs

~/Documents/Teaching/ENVS543/

```
> library(ggplot2)
> ggplot( iris ) +
+ geom_boxplot( aes(Species, Petal.Width) ) +
+ theme_minimal() +
+ xlab("Species of Iris") +
+ ylab("Petal Width")
>
```

Files | Plots | Packages | Help | Viewer

Zoom | Export



Persistent R Scripts

Bootcamp_Chronology.R

Source on Save | Run | Source

```
data <- read.csv( text = "Language,Year,Label
                  S,1975,Bell Laboratories Version 1.0
                  S,1980,Release Outside Bell Labs
                  S,1984,Education & Commercial Licensing
                  S,1988,Statiscal Software Inc.
                  S,1993,Mathsoft Buys S
                  S,1995,Version 3.3
                  S,1996,Version 3.4
                  S,1997,Version 4.0
                  S,1999,Version 5.0
                  S,2000,Version 6.0
                  S,2002,Free Student Version
                  S,2005,Version 7.0
                  S,2007,Version 8.0
                  S,2015,TIBCO
                  R,1995,Unstable
                  R,1997,Version 0.49
                  R,1999,Version 0.65 CRAN & Packages
                  R,2000,Version 1.0
                  R,2001,Version 1.4 S4
                  R,2004,Version 2.0
                  R,2013,Version 3.0
                  R,2017,Version 3.4
                  R,2018,Version 3.5")

data$Y <- c( 0.1, 0.2, 0.3, 0.3, 0.5, 0.55, 0.5, 0.45, 0.5, 0.4, 0.3, 0.2, 0.1, 0.1,
             -0.3, -0.3, -0.3, -0.25, -0.2, rep(-0.3, times=3), 0.25 )
data$ymax <- c( data$Y[1:14], rep(0,9))
data$ymin <- c( rep(0,14), data$Y[15:23])
data$angle <- c( rep(45,14), rep(-43,9) )
data$hjust <- c( rep(0,14), rep(0, 9))
ggplot( data, aes(Year,Y,label=Label) ) +
  geom_linerange( aes(ymin=ymin,ymax=ymax) ) +
  geom_text( aes( color=Language, angle=angle, hjust=hjust), vjust=0, size=3) +
  geom_hline(yintercept=0,color = "black", size=0.3) +
  geom_point( aes(y=0, color=Language), size=3) +
  scale_color_manual( values=c("#0070C0", "#00B050") ) +
  theme_classic() + ylim(c(-1,1.4) ) + xlim( c(1975,2021) ) +
  theme( axis.line.y=element_blank(),
                axis.text.y=element_blank(),
                axis.title.x=element_blank(),
                axis.title.y=element_blank(),
                axis.ticks.y=element_blank(),
                legend.position = "none" ) -> p

print(p)
```

46:9 (Top Level) R Script

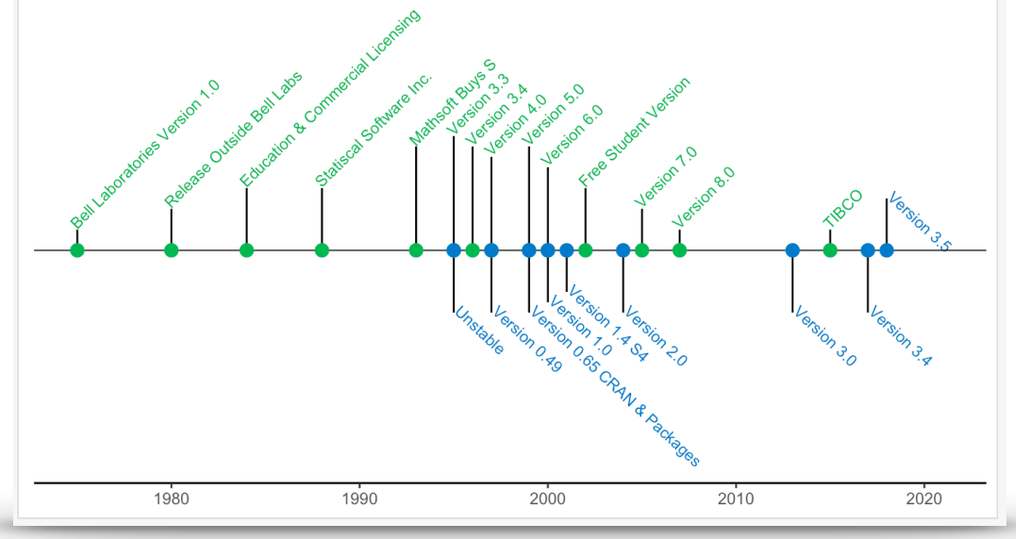
Console | Terminal | R Markdown | Jobs

~/Documents/Teaching/ENVS543/

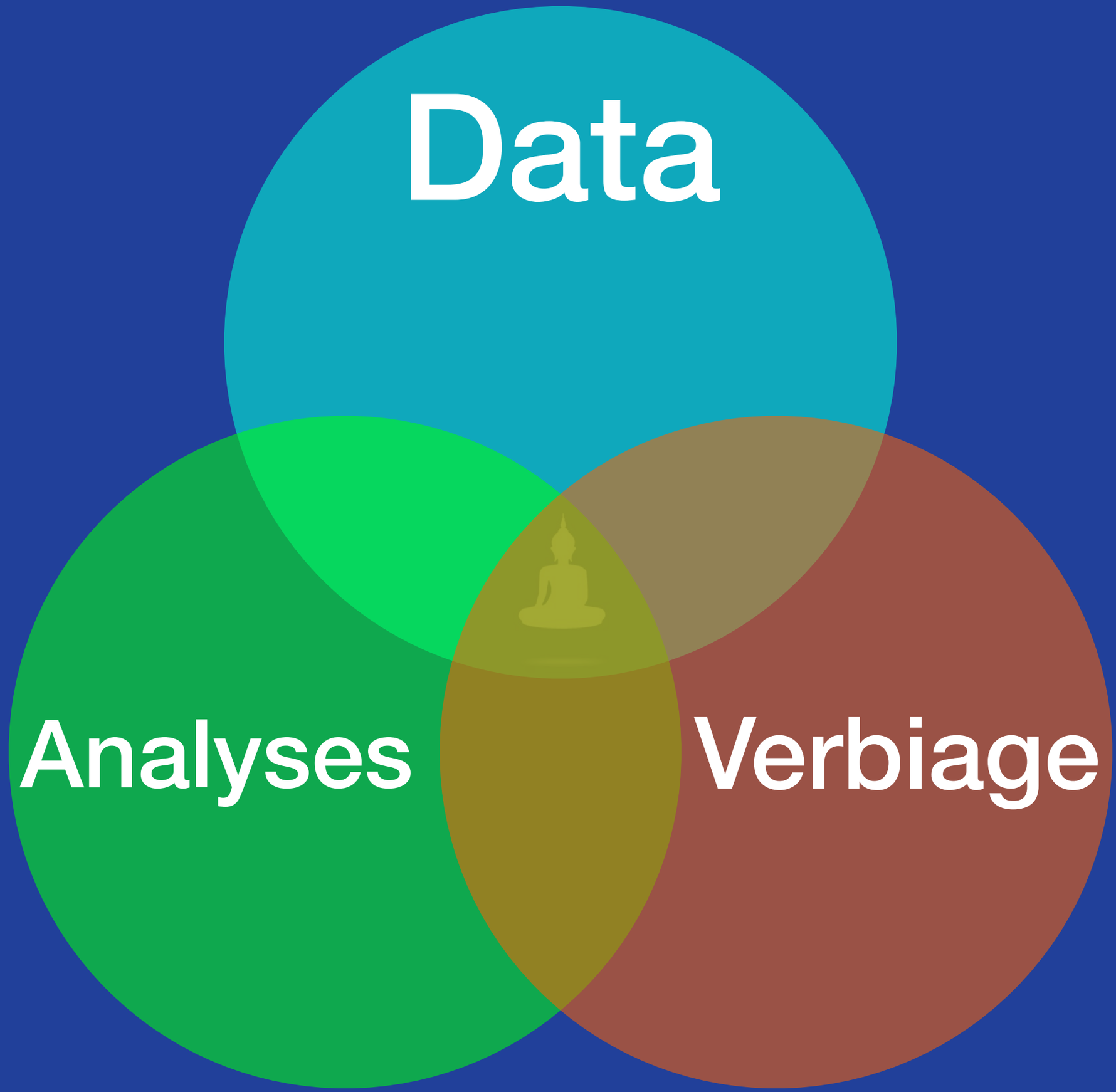
```
> source('~/Documents/Teaching/ENVS543/Content/Bootcamp/Bootcamp_Chronology.R')
>
```

Files | Plots | Packages | Help | Viewer

Zoom | Export | Publish



Reproducible Research



Data

Actual Interactive Table
In RStudio

Show 10 entries

Search:

	Site	Longitude	Latitude	Males	Females	Suitability
1	12	-112.66552	27.18232	24	21	0.35190499
2	153	-110.46236	24.13389	35	41	0.73248702
3	157	-110.096	24.0195	26	30	0.88102901
4	159	-113.31609	27.52944	22	15	0.18796501
5	160	-112.52959	27.40498	48	36	0.36519101
6	161	-112.986	27.0367	64	63	0.27910501
7	162	-112.408	27.2028	57	41	0.6136198
8	163	-110.951	24.2115	21	21	0.43287301
9	166	-112.08062	25.91409	19	26	0.26730299
10	168	-111.21563	25.55757	28	25	0.496464998

Showing 1 to 10 of 29 entries

Previous 1 2 3 Next

Data + Narrative

Show 10 entries

	Site	Longitude	Latitude	Males
1	12	-112.66552	27.18232	24
2	153	-110.46236	24.13389	35
3	157	-110.096	24.0195	26
4	159	-113.31609	27.52944	22
5	160	-112.52959	27.40498	48
6	161	-112.986	27.0367	64
7	162	-112.408	27.2028	57
8	163	-110.951	24.2115	21
9	166	-112.08062	25.91409	19
10	168	-111.21563	25.55757	28

Showing 1 to 10 of 29 entries

The spatial distribution of $N = 1408$ individuals were sampled from $K = 29$ sampling sites covered the entire known extent for *Araptus attenuatus* (Figure 1). On average the sex ratio (Females:Males) within each of the sampling locales varied considerably, ranging from a low of 0.455 to 1.684.

Evidence of differential sex-biased dispersal is revealed by differential structure in biparental and uniparentally inherited genetic markers. The expectation under equal dispersal betweeen sezes for maternally inherited mitochondrial markers ($F_{ST;mt}$) and biparent nuclear markers ($F_{ST;nuc}$) are (see also Hartl & Clark 1997; Wang 1997 for another formulation):

$$F_{ST;mt} = \frac{8F_{ST;nuc}}{1 + 7F_{ST;nuc}}$$

Actual Manuscript Text
Written In RStudio

Data + Narrative + Figures

Show 10 entries — Search:

	Site	Longitude	Latitude	Males	Females	Suitability
1	12	-112.66552	27.18232	24	21	0.35190499
2	153	-110.46236				
3	157	-110.096				
4	159	-113.31609				
5	160	-112.52959				
6	161	-112.986				
7	162	-112.408				
8	163	-110.951				
9	166	-112.08062				
10	168	-111.21563				

Showing 1 to 10 of 29 entries

Sex ratio bias is not spatially homogeneous and has been shown to be correlated with the location within the landscape of individual populations. For example, several studies have shown that the relative density of females bears declines precipitously as one samples from the core to the periphery of the species range (e.g., Alt 1978; Rogers 1987; Swartz & Franzmann 1992). This spatial variation in sex ratio is maintained by increased dispersal rates for males over that of females (e.g., Swenson et al. 1998).

From a genetic perspective, evidence of differential sex-biased dispersal can be quantified by comparing the observed level of genetic differentiation among uniparentally inherited markers (say from the mitochondrial genome) to that observed in bi-parentally inherited markers. Under an island model of migration if the rate of gene flow is the same among both sexes then mitochondrial genetic differentiation (denoted FST;mt) and nuclear differentiation (FST; nuc) should be related as (see also Hartl & Clark 1997; Wang 1997 for another formulation):

$$F_{ST;mt} = \frac{8F_{ST;nuc}}{1 + 7F_{ST;nuc}}$$

Deviations in sex ratio can have a significant influence on the level of genetic diversity maintained within populations mainly through its influence on the genetic effective population size. The genetic effective number of individuals within a population is maximized with equal numbers of males and females and is minimized when one sex predominates (e.g., Hedrick 2009). Reductions in the genetic effective number of individuals can increase the potential of genetic drift to reduce overall genetic diversity, a spatial pattern observed repeatedly in natural populations (e.g., Wisely et al. 2004, Schwartz et al. 2003, Kark et al. 2008 among others).

The spatial distribution of $N = 1408$ individuals were sampled from $K = 29$ sampling sites covered the entire known extent for *Araptus attenuatus* (Figure 1). On average the sex ratio (Females:Males) within each of the sampling locales varied considerably, ranging from a low of 0.455 to 1.684.

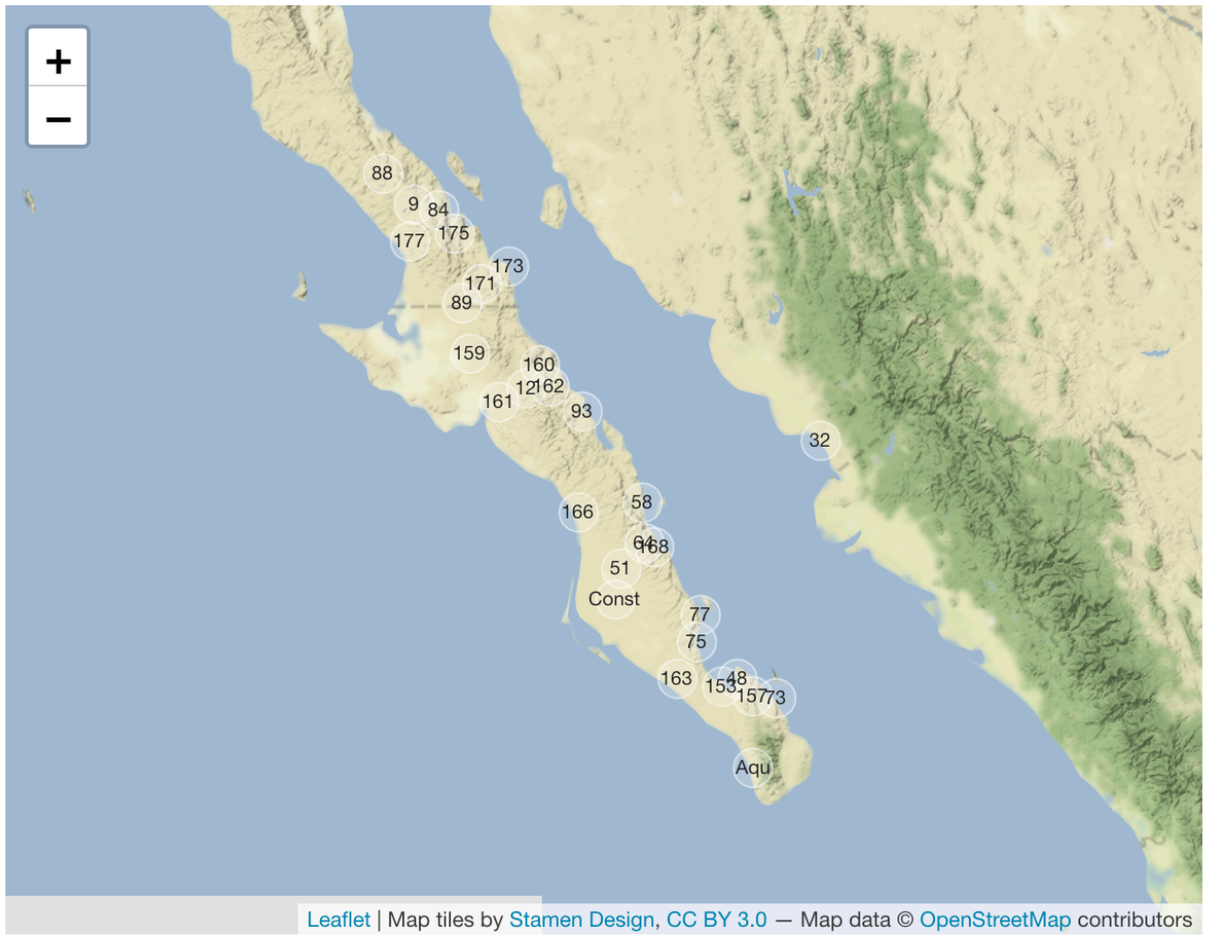


Figure 1: The spatial distribution of sample locations for the Sonoran Desert bark beetle.

Evidence of differential sex-biased dispersal is revealed by differential structure in biparental and uniparentally inherited genetic markers. The expectation under equal dispersal betweeen sezes for maternally inherited mitochondrial markers ($F_{ST;mt}$) and biparent nuclear markers ($F_{ST;nuc}$) are (see also Hartl & Clark 1997; Wang 1997 for another formulation):

$$F_{ST;mt} = \frac{8F_{ST;nuc}}{1 + 7F_{ST;nuc}}$$

Data + Narrative + Figures + Analyses

Show 10 entries

	Site	Longitude	Latitude	Males	Fe
1	12	-112.66552	27.18232	24	
2	153	-110.46236			
3	157	-110.096			
4	159	-113.31609			
5	160	-112.52959			
6	161	-112.986			
7	162	-112.408			
8	163	-110.951			
9	166	-112.08062			
10	168	-111.21563			

Showing 1 to 10 of 29 entries

Sex ratio bias is not spatially homogeneo
of individual populations. For example, se
precipitously as one samples from the co
Franzmann 1992). This spatial variation in
females (e.g., Swenson et al. 1998).

From a genetic perspective, evidence of
level of genetic differentiation among unip
observed in bi-parentally inherited marke
both sexes then mitochondrial genetic dif
related as (see also Hartl & Clark 1997; W

Deviations in sex ratio can have a signific
mainly through its influence on the geneti
population is maximized with equal numb
Hedrick 2009). Reductions in the genetic
reduce overall genetic diversity, a spatial
Schwartz et al. 2003, Kark et al. 2008 am

The spatial distribution of $N = 1408$ individuals were sampled
covered the entire known extent for *Araptus attenuatus* (Figure
(Females:Males) within each of the sampling locales varied cor
0.455 to 1.684.

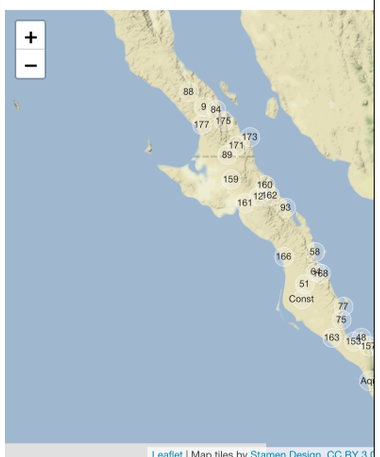


Figure 1: The spatial distribution of sample locations for t

Evidence of differential sex-biased dispersal is revealed by diff
uniparentally inherited genetic markers. The expectation under
maternally inherited mitochondrial markers ($F_{ST;mt}$) and bipare
also Hartl & Clark 1997; Wang 1997 for another formulation):

$$F_{ST;mt} = \frac{8F_{ST;nuc}}{1 + 7F_{ST;nuc}}$$

Results

Female beetles seem to be selecting locales in more suitable habitats (Figure 2), which is consistent with a Gaussian log normal distribution (GLM; intercept = 0.836; coef =-0.216; P =0.029). This suggests that males who establish brood chambers at the margins of the species niche are actively selected against by maturing females.

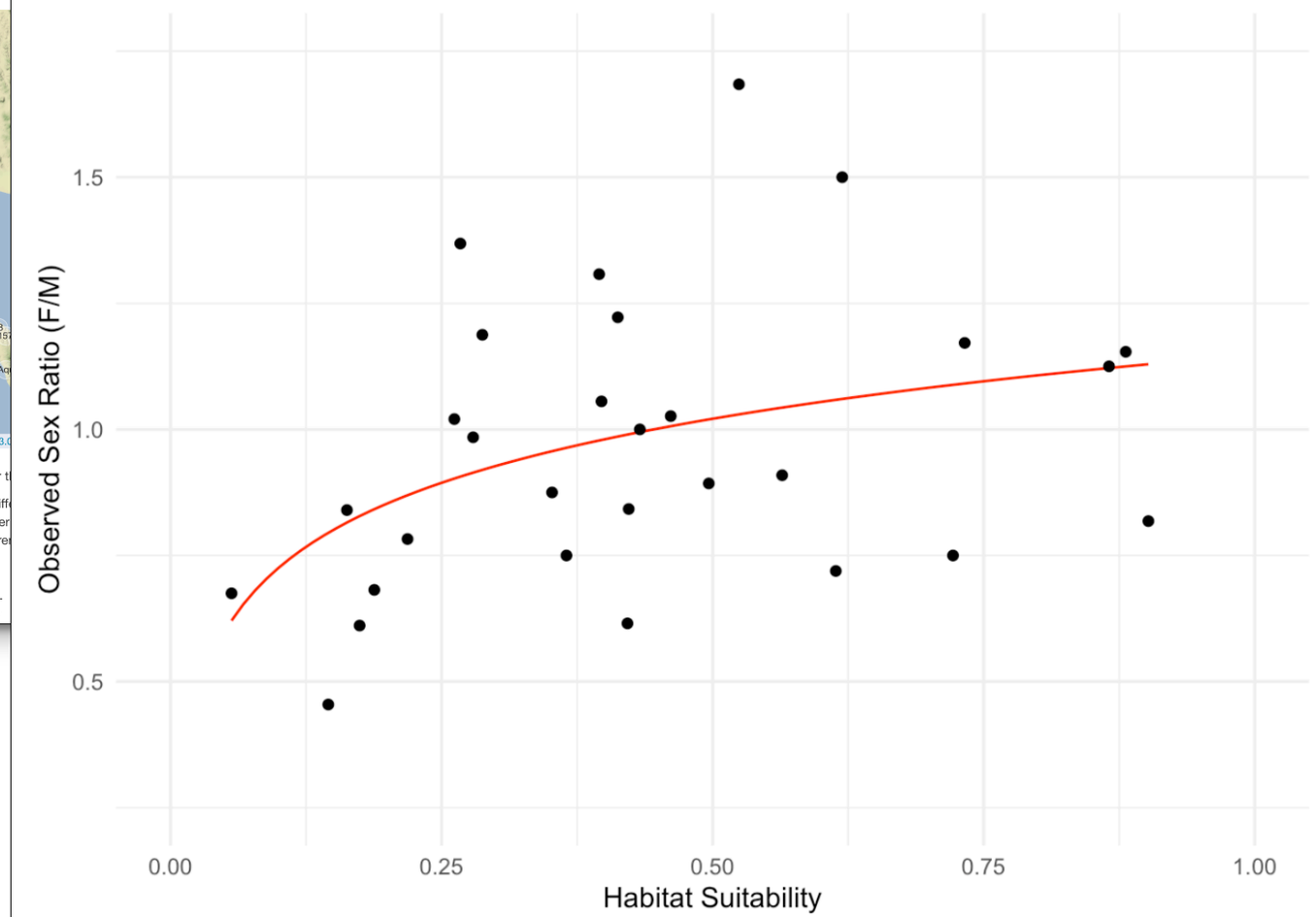


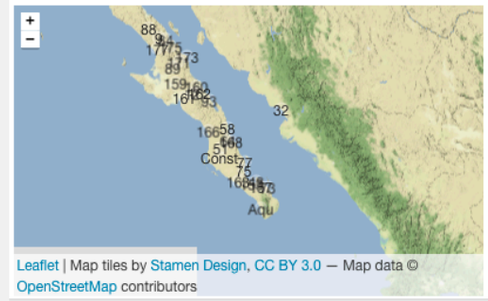
Figure 2: Variation in the observed sex ratio of *A. attenuatus* locales as a function of habitat suitability as determined by MAXENT Niche modeling of host plant distributions.

R & Quarto

- Data + Narrative + Figures + Analyses in one document.
- Raw `text` format
  - Longevity
  - Collaboration
  - Reproducibility
- Dynamic data visualization tools

````
  addCircleMarkers(weight=1, fill=TRUE, color='#ff7f7f') %>%
  addLabelOnlyMarkers(label=~as.character(Site),
                      labelOptions = labelOptions(noHide=TRUE,
                                                  direction='center',
                                                  textOnly = TRUE))


```
````




````
Evidence of differential sex-biased dispersal is revealed by
differential structure in biparental and uniparentally inherited
genetic markers.  The expectation under equal dispersal betweeen
sezes for maternally inherited mitochondrial markers ($F_{ST;mt}$)
and biparent nuclear markers ($F_{ST; nuc}$) are (see also Hartl &
Clark 1997; Wang 1997 for another formulation):

\[
F_{ST; mt}= \frac{8 F_{ST;nuc}}{1 + 7 F_{ST; nuc} }
\]


```{r echo=TRUE}
library(leaflet)
library(leaflet.minicharts)

leaflet() %>%
  addProviderTiles( providers$Esri.WorldTopoMap) %>%
  addMinicharts(
    araptus$Longitude, araptus$Latitude,
    type = "pie",
    chartdata = araptus[, c("Males","Females")],
    width = araptus$Suitability*30
    )


```


## Results

```{r}
araptus %>%
  mutate( SexRatio = Females / Males ) -> araptus

model <- glm( SexRatio ~ log(Suitability),
              data = araptus,
              family = Gamma() )
# summary(model)
```


Call:
glm(formula = SexRatio ~ log(Suitability), family = Gamma(),
    data = araptus)

Deviance Residuals:
     Min        1Q    Median        3Q       Max
-0.51536  -0.19664  -0.01675   0.09144   0.54147

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)      0.83650    0.09529   8.778 2.15e-09 ***
log(Suitability) -0.21558    0.09333  -2.310   0.0288 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for Gamma family taken to be 0.07508152)

    Null deviance: 2.4077  on 28  degrees of freedom
Residual deviance: 1.9864  on 27  degrees of freedom
AIC: 6.4873

Number of Fisher Scoring iterations: 4


Female beetles seem to be selecting locales in more suitable habitats
(Figure 2), which is consistent with a Gaussian log normal
distribution (GLM; intercept = `r summary(model)$coefficients[1,1]`;
coef =`r summary(model)$coefficients[2,1]`; P =`r
summary(model)$coefficients[2,4]`).  This suggests that males who
establish brood chambers at the margins of the species niche are
actively selected against by maturing females.

```{r fig.cap="*Figure 2:* Variation in the observed sex ratio of *A.
attenuatus* locales as a function of habitat suitability as
determined by MAXENT Niche modeling of host plant distributions.",
````

Built-In Graphics

Configuration of the graph is done largely by adding more content to the `plot()` function.

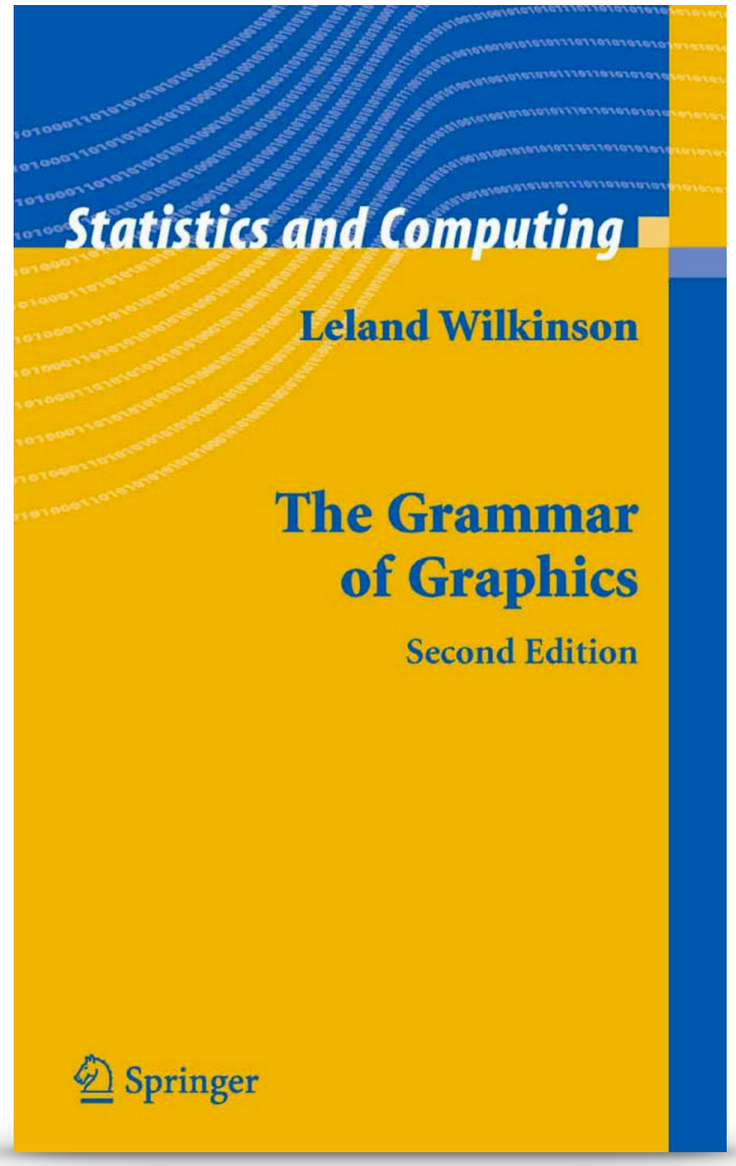
Usage

```
## Default S3 method:
plot(x, y = NULL, type = "p",  xlim = NULL, ylim = NULL,
     log = "", main = NULL, sub = NULL, xlab = NULL, ylab = NULL,
     ann = par("ann"), axes = TRUE, frame.plot = axes,
     panel.first = NULL, panel.last = NULL, asp = NA, ...)
```

Graphics

Grammar of Graphics

- Data
- Aesthetics (color, size, variables)
- Geometries (points, lines, histograms)
- Partitions
- Transformation (polar, cartesian, projection)
- Thematic display



Tidyverse

R packages for data science

The tidyverse is an opinionated collection of R packages designed for data science. All packages share an underlying philosophy and common APIs.

Install the complete tidyverse with:

```
install.packages("tidyverse")
```

Example Data

```
summary(airquality)
```

```
     Ozone           Solar.R          Wind             Temp           Month            Day      
 Min.   :  1.00   Min.   :  7.0   Min.   : 1.700   Min.   :56.00   Min.   :5.000   Min.   : 1.0  
 1st Qu.: 18.00   1st Qu.:115.8   1st Qu.: 7.400   1st Qu.:72.00   1st Qu.:6.000   1st Qu.: 8.0  
 Median : 31.50   Median :205.0   Median : 9.700   Median :79.00   Median :7.000   Median :16.0  
 Mean   : 42.13   Mean   :185.9   Mean   : 9.958   Mean   :77.88   Mean   :6.993   Mean   :15.8  
 3rd Qu.: 63.25   3rd Qu.:258.8   3rd Qu.:11.500   3rd Qu.:85.00   3rd Qu.:8.000   3rd Qu.:23.0  
 Max.   :168.00   Max.   :334.0   Max.   :20.700   Max.   :97.00   Max.   :9.000   Max.   :31.0  
 NA's   :37       NA's   :7                                                                      
```

Missing Data For Some

Not Others

Wonky Dates (Year?)

Data Munging (for sanity)

```
summary(airquality)
```

```
     Ozone           Solar.R          Wind             Temp           Month            Day      
 Min.   :  1.00   Min.   :  7.0   Min.   : 1.700   Min.   :56.00   Min.   :5.000   Min.   : 1.0  
 1st Qu.: 18.00   1st Qu.:115.8   1st Qu.: 7.400   1st Qu.:72.00   1st Qu.:6.000   1st Qu.: 8.0  
 Median : 31.50   Median :205.0   Median : 9.700   Median :79.00   Median :7.000   Median :16.0  
 Mean   : 42.13   Mean   :185.9   Mean   : 9.958   Mean   :77.88   Mean   :6.993   Mean   :15.8  
 3rd Qu.: 63.25   3rd Qu.:258.8   3rd Qu.:11.500   3rd Qu.:85.00   3rd Qu.:8.000   3rd Qu.:23.0  
 Max.   :168.00   Max.   :334.0   Max.   :20.700   Max.   :97.00   Max.   :9.000   Max.   :31.0  
 NA's   :37       NA's   :7
```

```
airquality %>%
  mutate( Month = factor( Month, ordered = TRUE),
          Day = factor( Day, ordered = TRUE ) ) -> airquality
summary( airquality)
```

```
     Ozone           Solar.R          Wind             Temp           Month       Day      
 Min.   :  1.00   Min.   :  7.0   Min.   : 1.700   Min.   :56.00   5:31   1      :  5  
 1st Qu.: 18.00   1st Qu.:115.8   1st Qu.: 7.400   1st Qu.:72.00   6:30   2      :  5  
 Median : 31.50   Median :205.0   Median : 9.700   Median :79.00   7:31   3      :  5  
 Mean   : 42.13   Mean   :185.9   Mean   : 9.958   Mean   :77.88   8:31   4      :  5  
 3rd Qu.: 63.25   3rd Qu.:258.8   3rd Qu.:11.500   3rd Qu.:85.00   9:30   5      :  5  
 Max.   :168.00   Max.   :334.0   Max.   :20.700   Max.   :97.00          6      :  5  
 NA's   :37       NA's   :7                                               (Other):123
```

Constructing a Graph

Literally Add These Components Together

- Data
- Aesthetics (color, size, variables)
- Geometries (points, lines, histograms)
- Partitions
- Transformation (polar, cartesian, projection)
- Thematic display

```
ggplot()  +  geom_XXXXX()   +  geom_YYYY() + coord_XXXX() + stat_XXXX()  + theme_XXX()
```

```
ggplot( data, aes(x=*,y=*)) + geom_***() + scale_x_***() + theme_*()
```

Aesthetics

Aesthetic Mapping `aes()`

```
aes( x = Temp, y = Wind )
```

```
Aesthetic mapping:
* `x` -> `Temp`
* `y` -> `Wind`
```

Maps variables in the **graph** to those in the **data**

- x, y, z
- shape, color, fill, grouping
- label,
- stat,
- binsize,
- alpha
- …

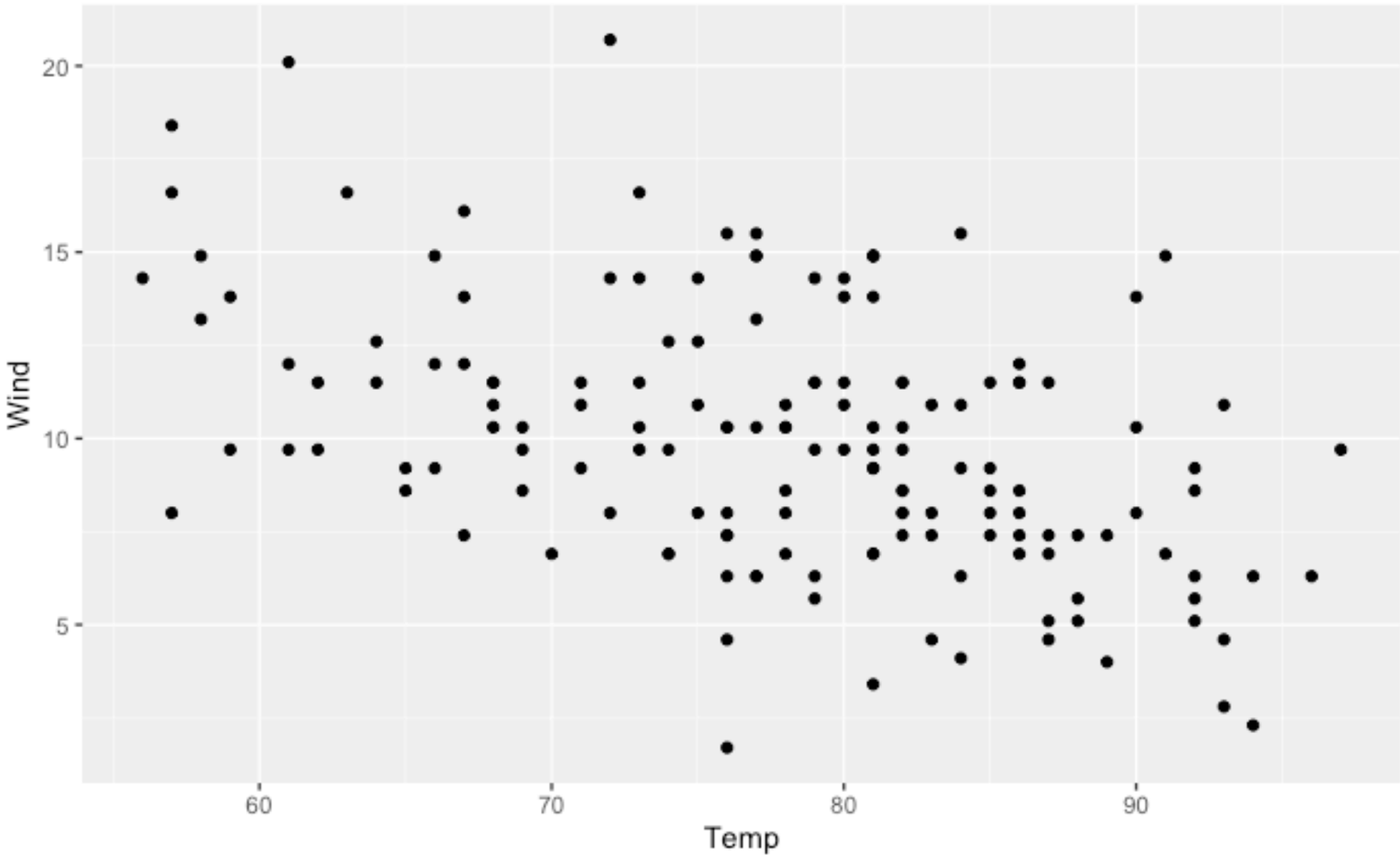
```
ggplot( data, aes(x=,y=)) + geom_***() + scale_x_***() + theme_*()
```

The Data Part

Need Data!

```
ggplot( airquality, aes( x = Temp, y = Wind) ) + geom_point()
```

- **Data Placement**
- **Aesthetics**
- **Options**
- **Order**



```
ggplot( data, aes(x=,y=)) + geom_***() + scale_x_***() + theme_*()
```

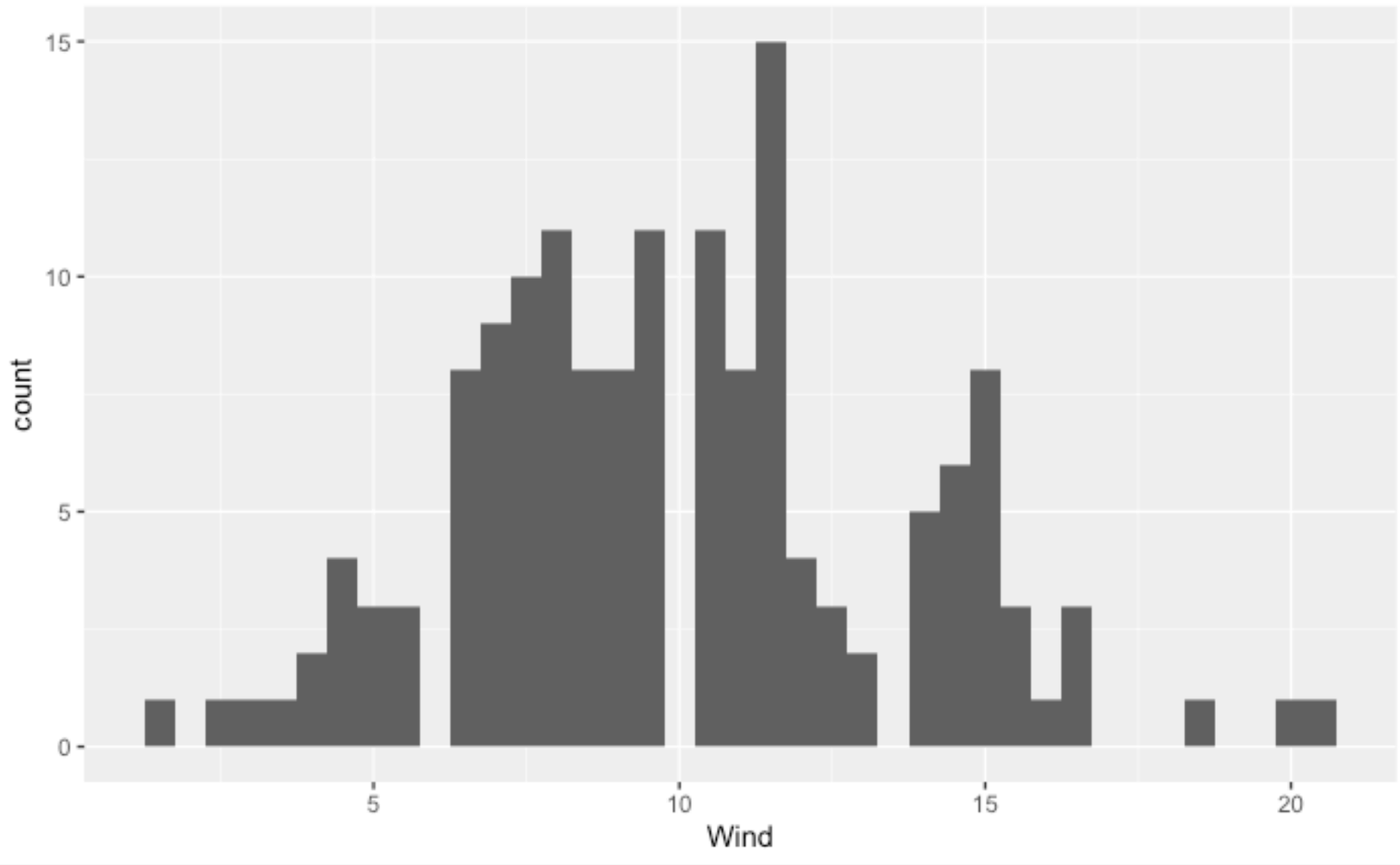
Geometries For 1 Variable

Histogram

```
ggplot( airquality, aes( Wind ) )
```

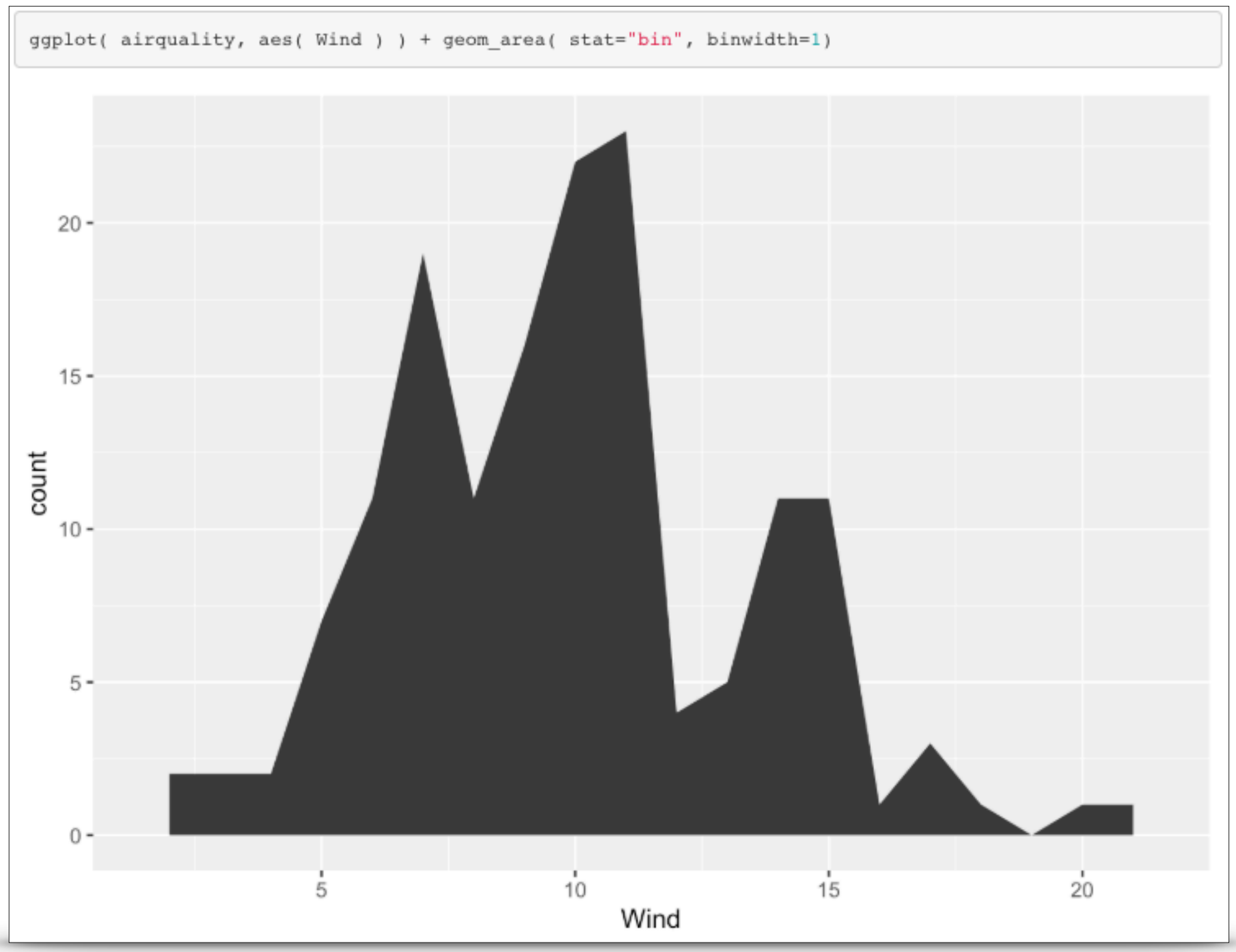
Histogram - Bins Specified

```
ggplot( airquality, aes( Wind ) ) + geom_histogram( binwidth=0.5)
```



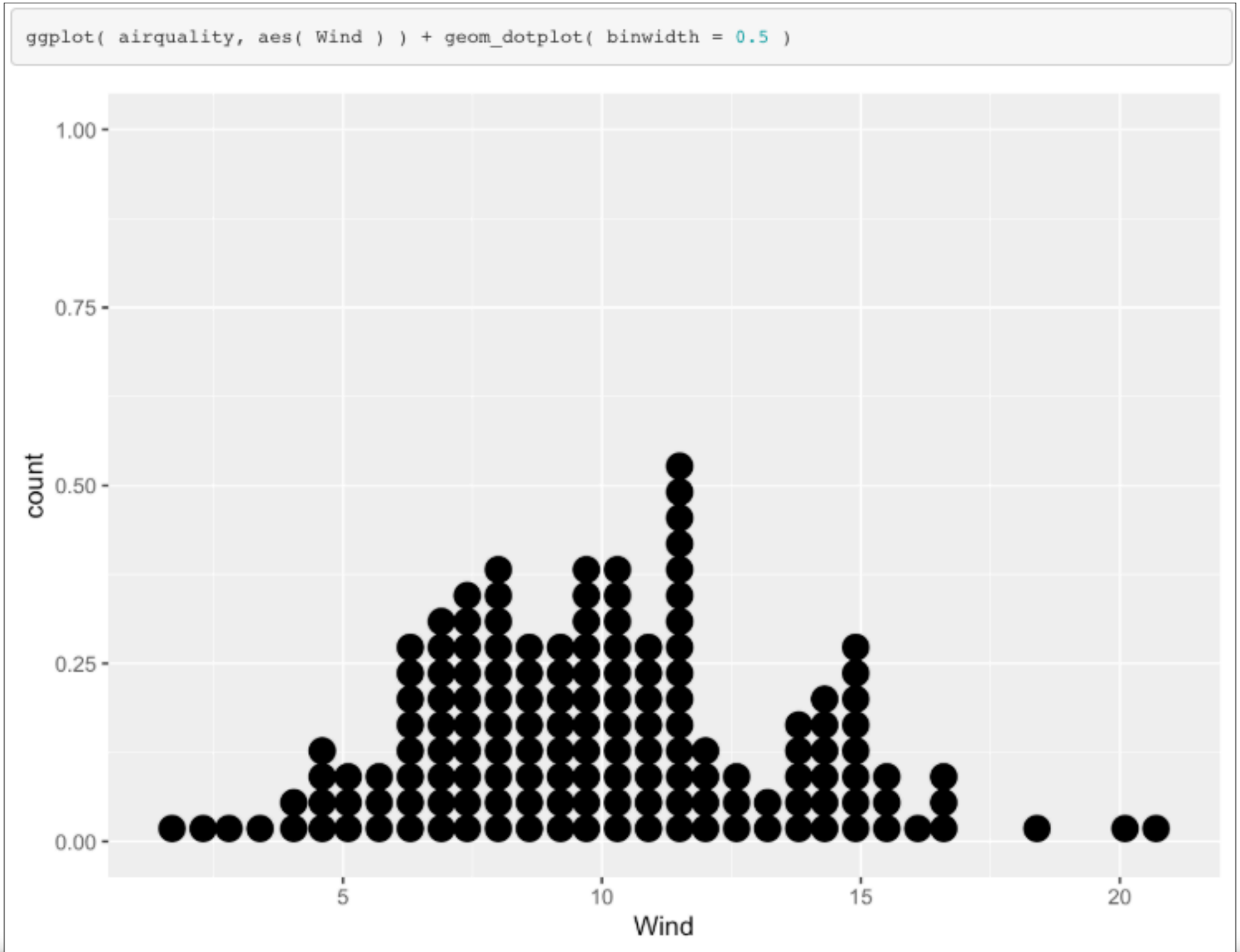
Area Plot

```
ggplot( airquality, aes( Wind ) ) + geom_area( stat="bin", binwidth=1)
```



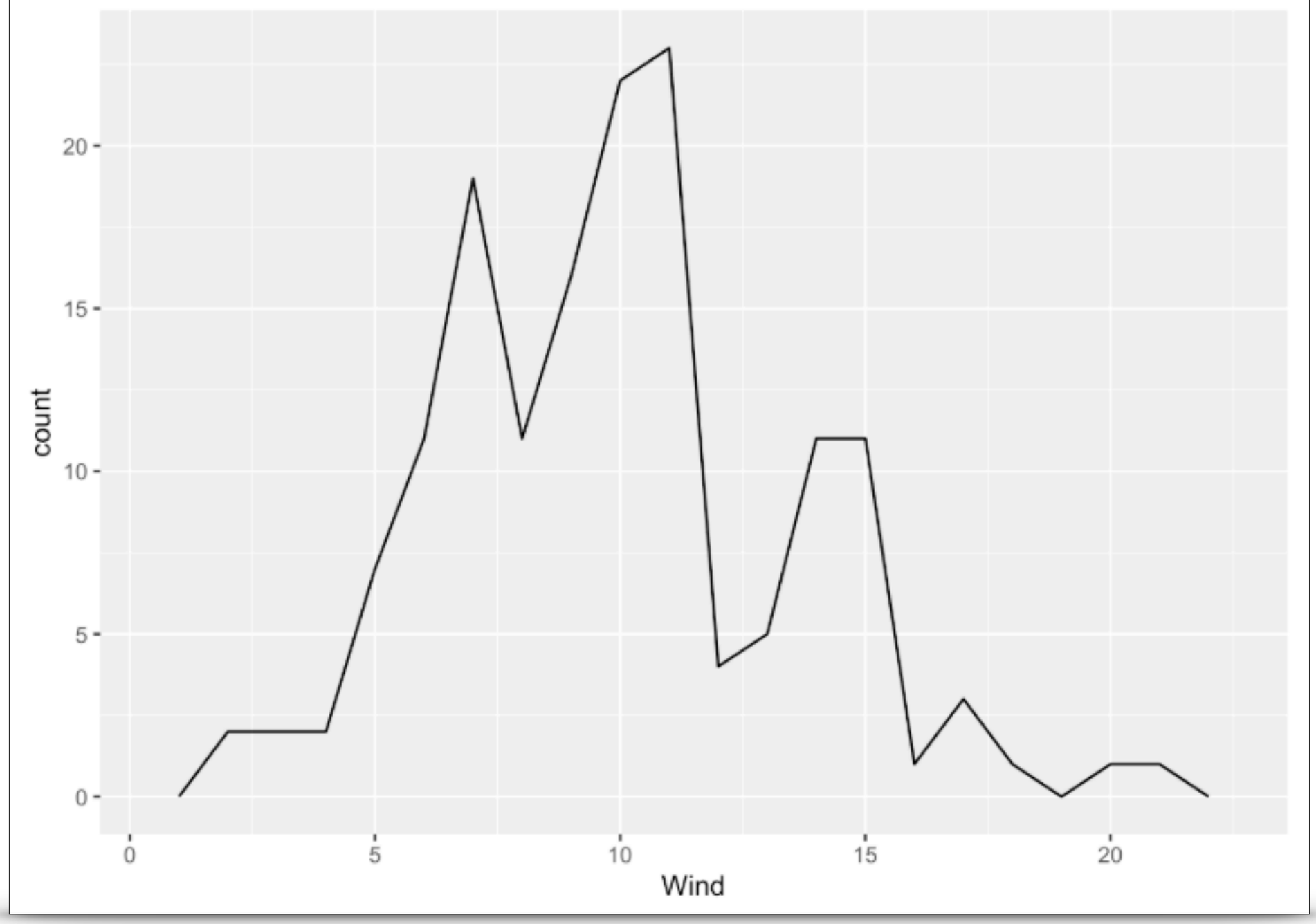
DotPlot

```
ggplot( airquality, aes( Wind ) ) + geom_dotplot( binwidth = 0.5 )
```

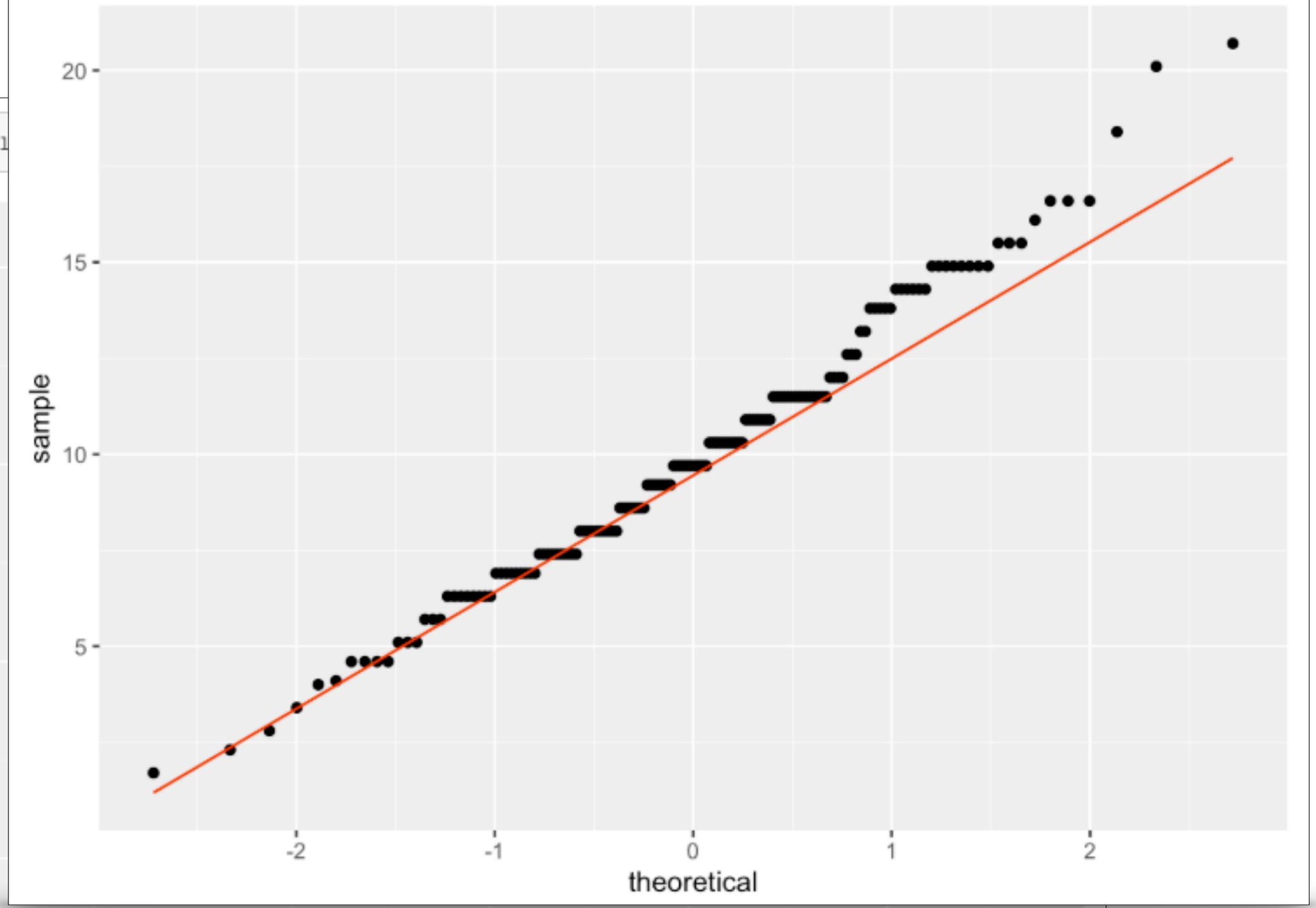


Frequency Polygon

```
ggplot( airquality, aes( Wind ) ) + geom_freqpoly( binwidth = 1.0 )
```



```
ggplot( airquality, aes( sample=Wind ) ) + geom_qq() + geom_qq_line( color="red")
```



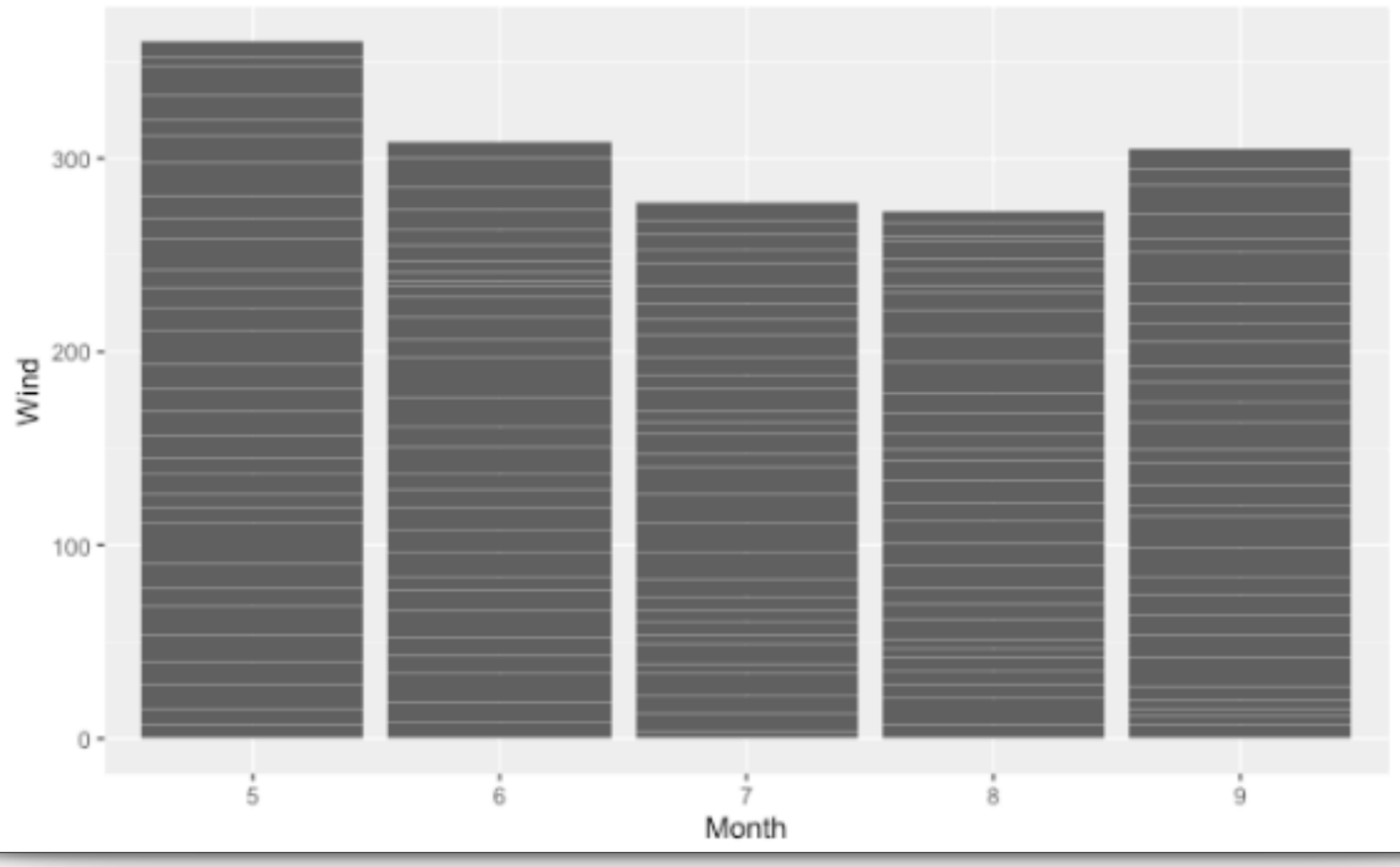

```
ggplot( airqual
```

```
ggplot( data, aes(x=,y=)) + geom_***() + scale_x_***() + theme_*()
```

Geometries For 2 Variables

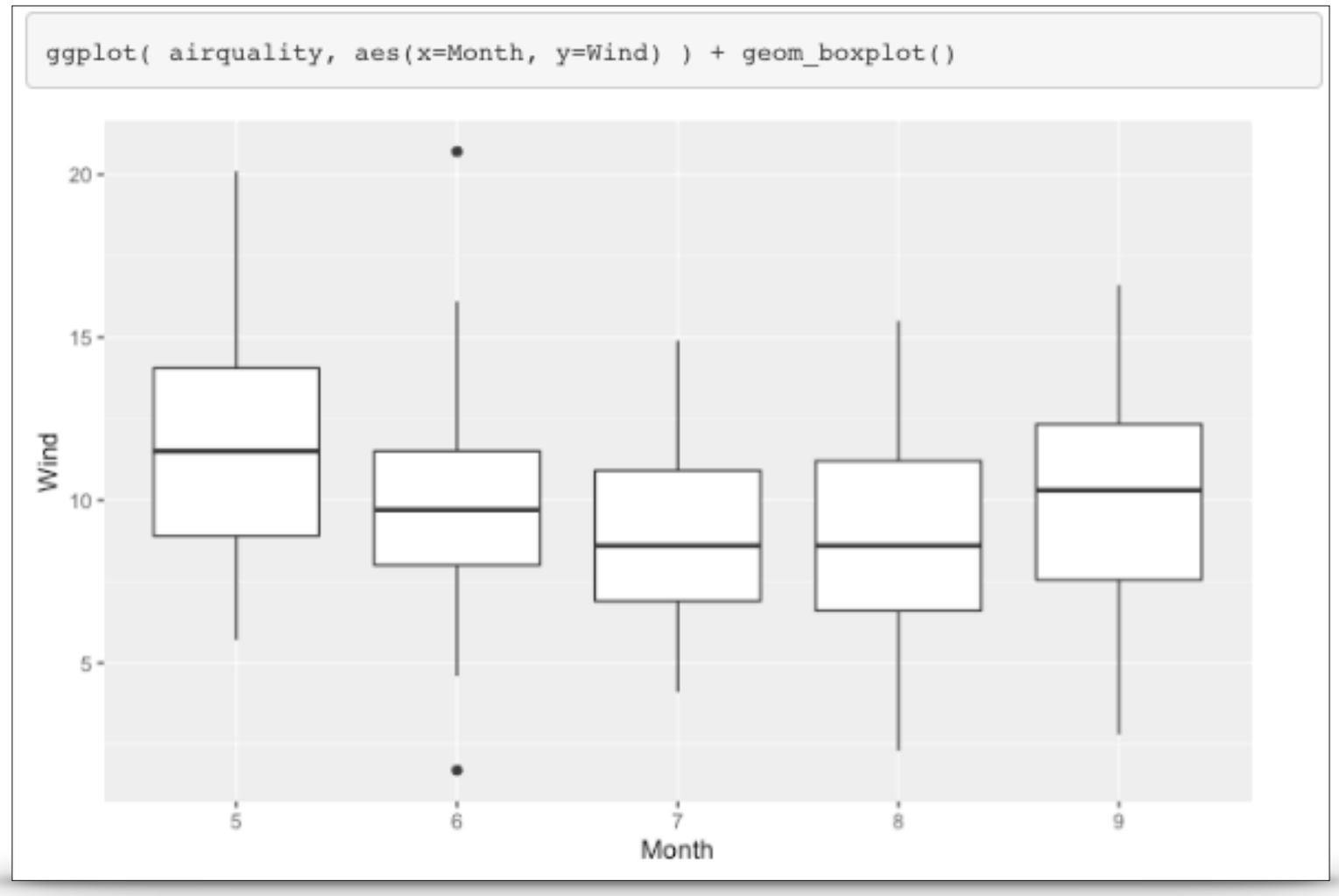
Columns

```
ggplot( airquality, aes(x=Month, y=Wind) ) + geom_col()
```



Boxplot

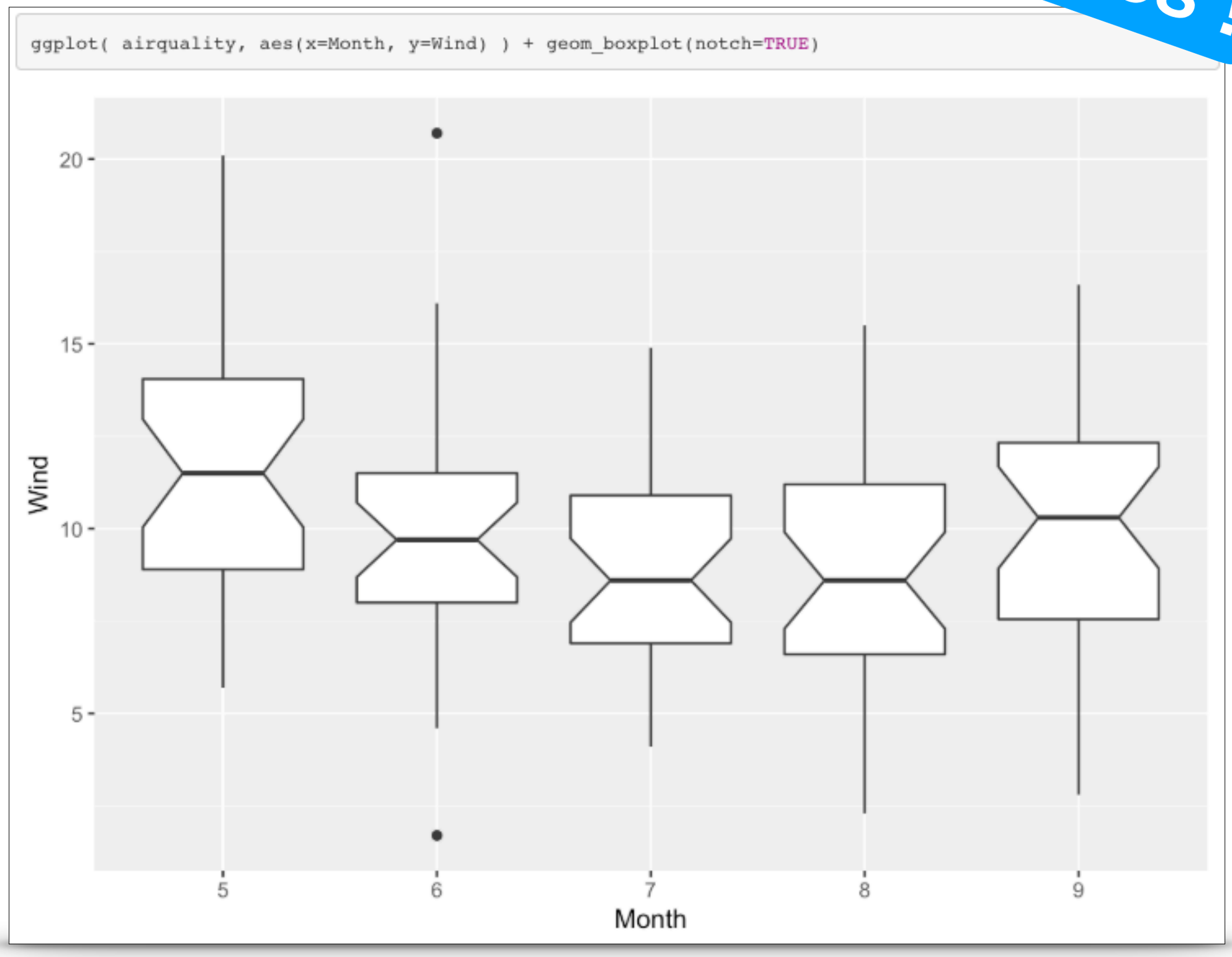
```
ggplot( airquality, aes(x=Month, y=Wind) ) + geom_boxplot()
```



Boxplot

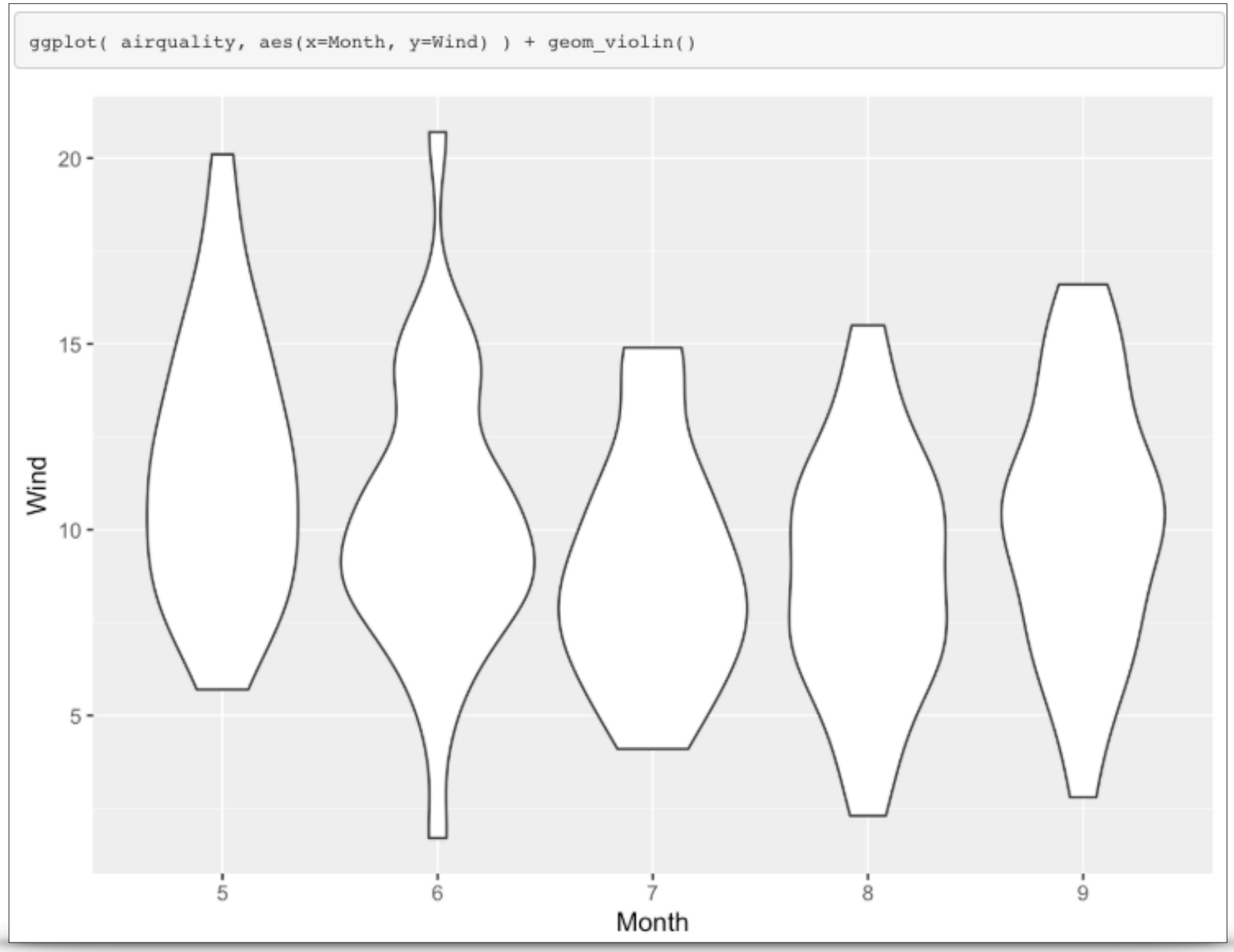
Now with Notches !

```
ggplot( airquality, aes(x=Month, y=Wind) ) + geom_boxplot(notch=TRUE)
```



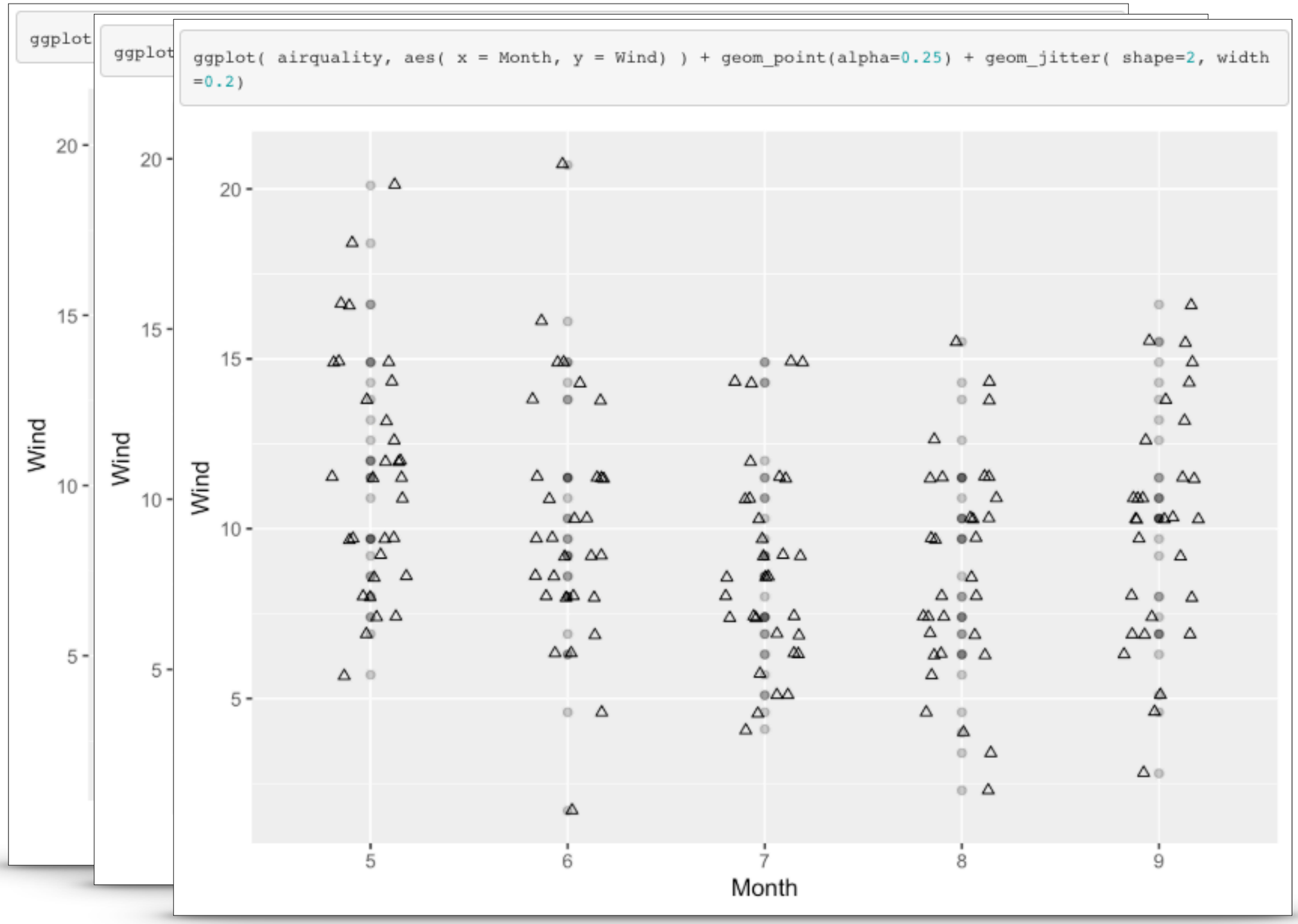
The Violin Plot

```
ggplot( airquality, aes(x=Month, y=Wind) ) + geom_violin()
```



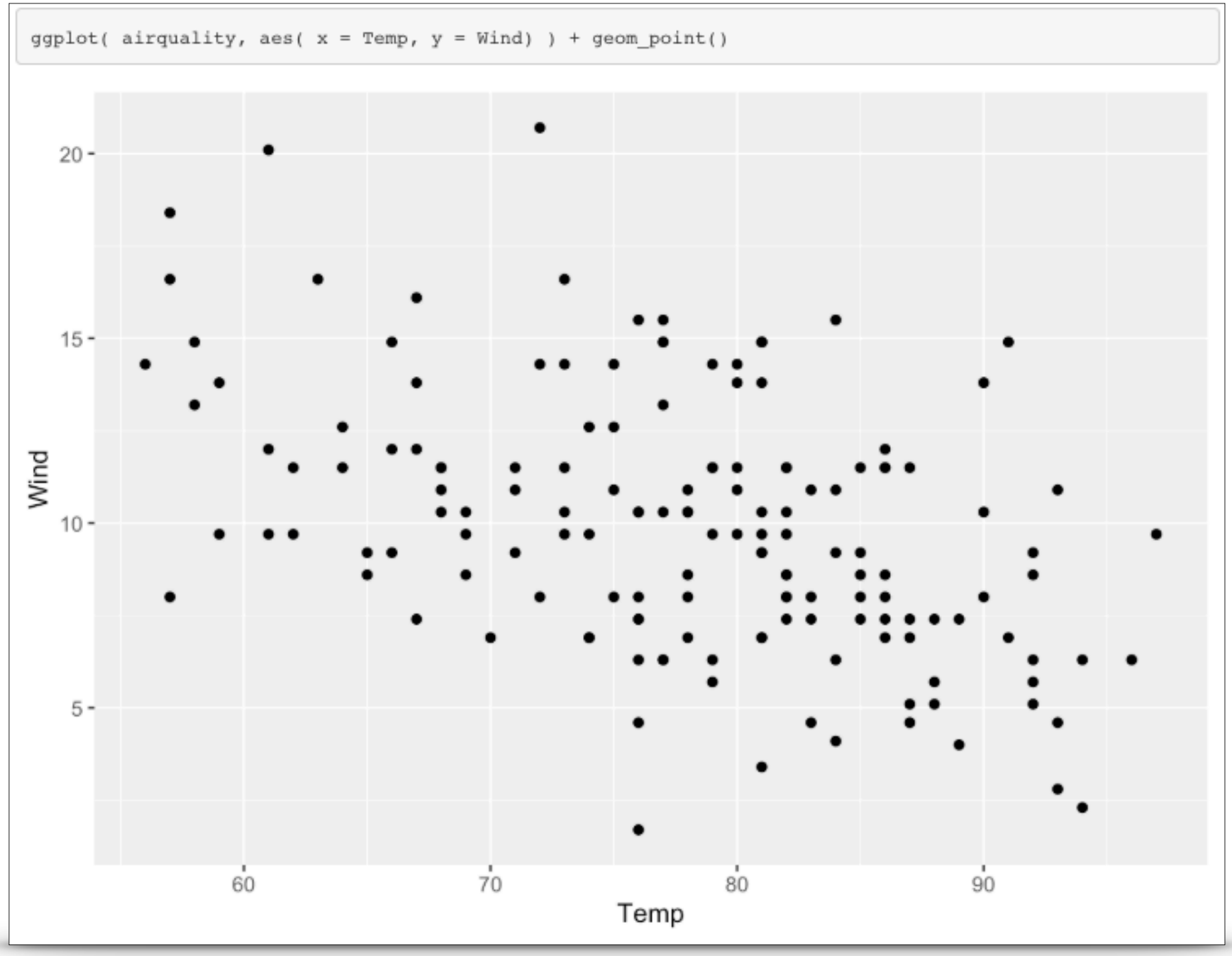
Points

```
ggplot( airquality, aes( x = Month, y = Wind) ) + geom_point(alpha=0.25) + geom_jitter( shape=2, width
=0.2)
```



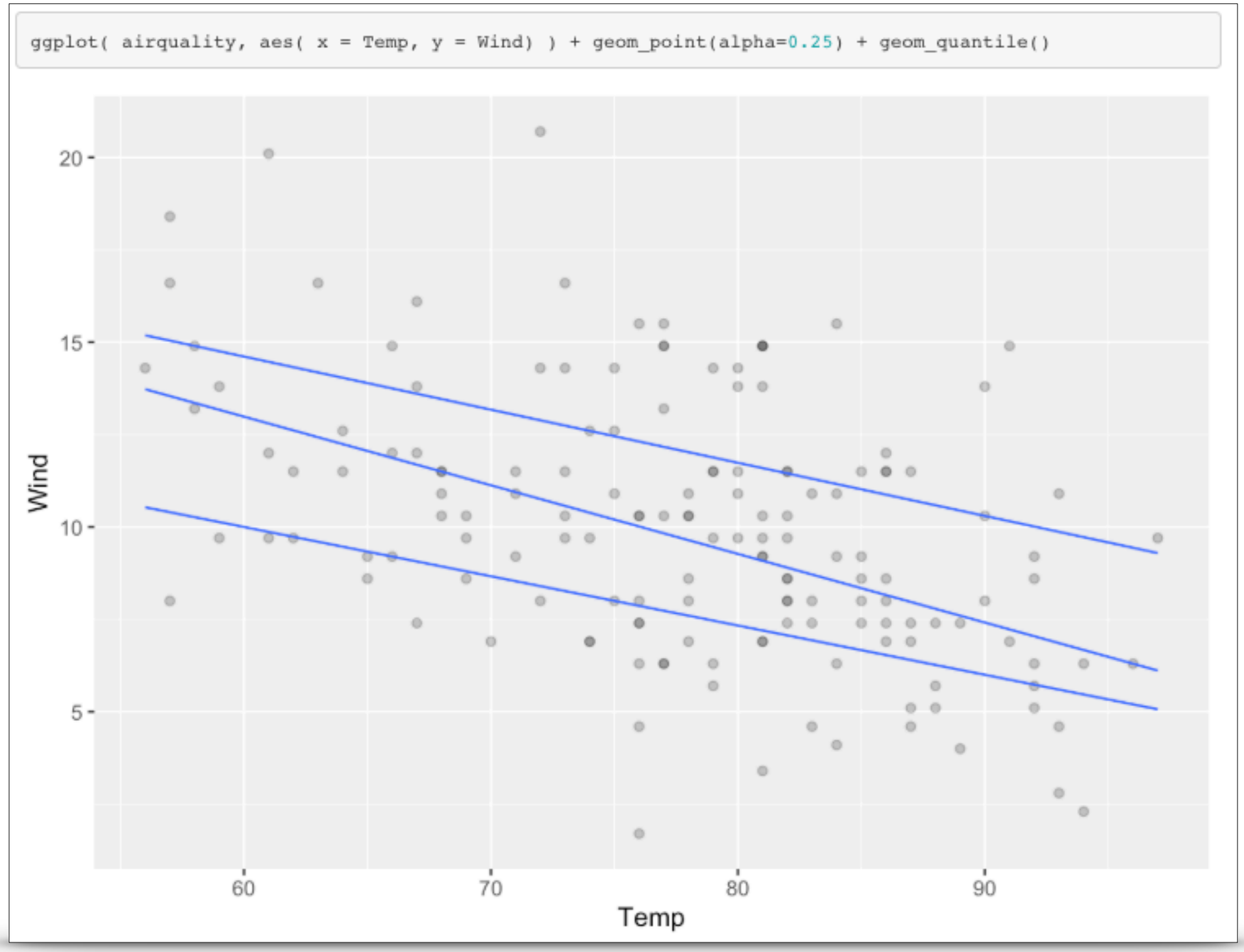
Points

```
ggplot( airquality, aes( x = Temp, y = Wind) ) + geom_point()
```



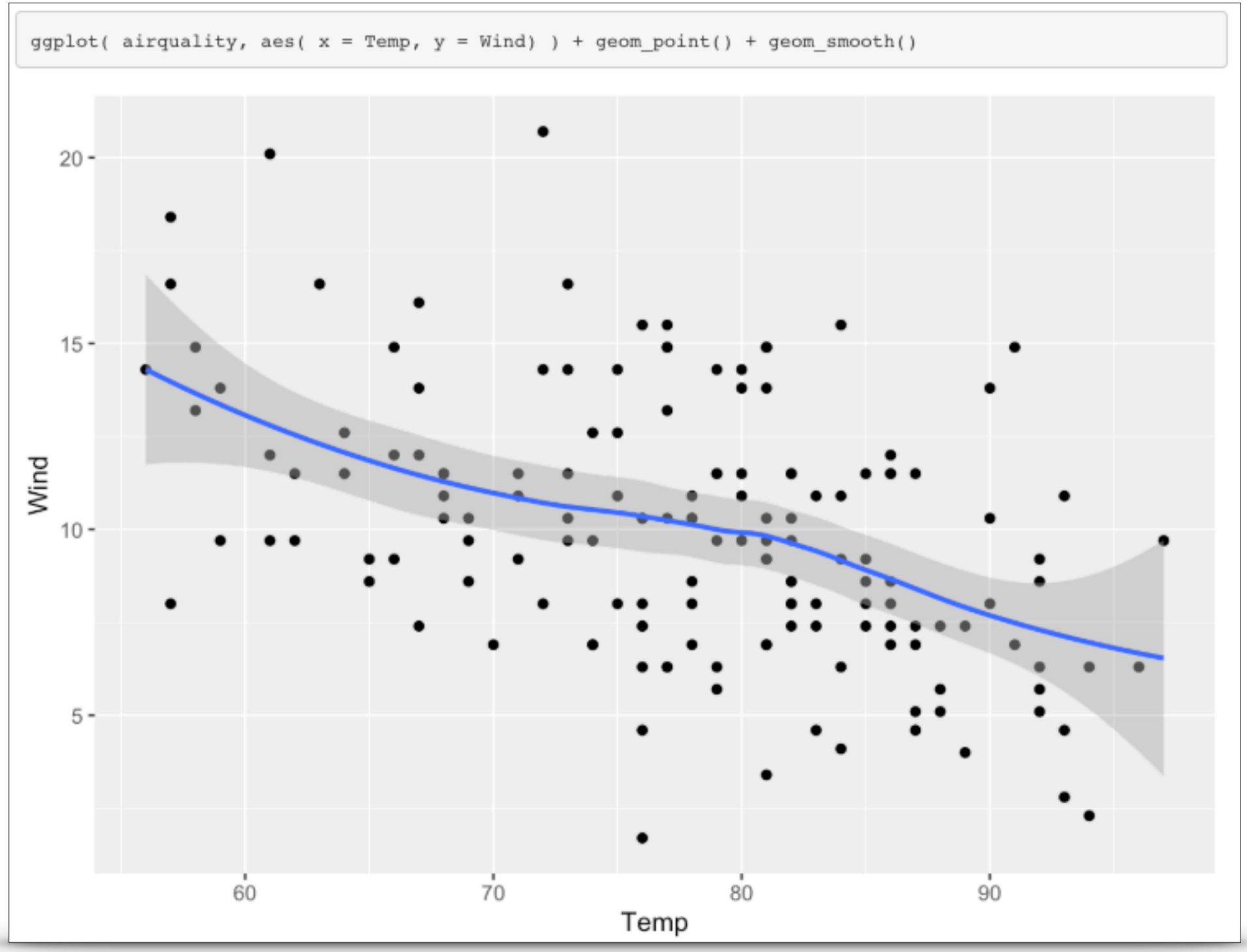
Quantiles over Points

```
ggplot( airquality, aes( x = Temp, y = Wind) ) + geom_point(alpha=0.25) + geom_quantile()
```



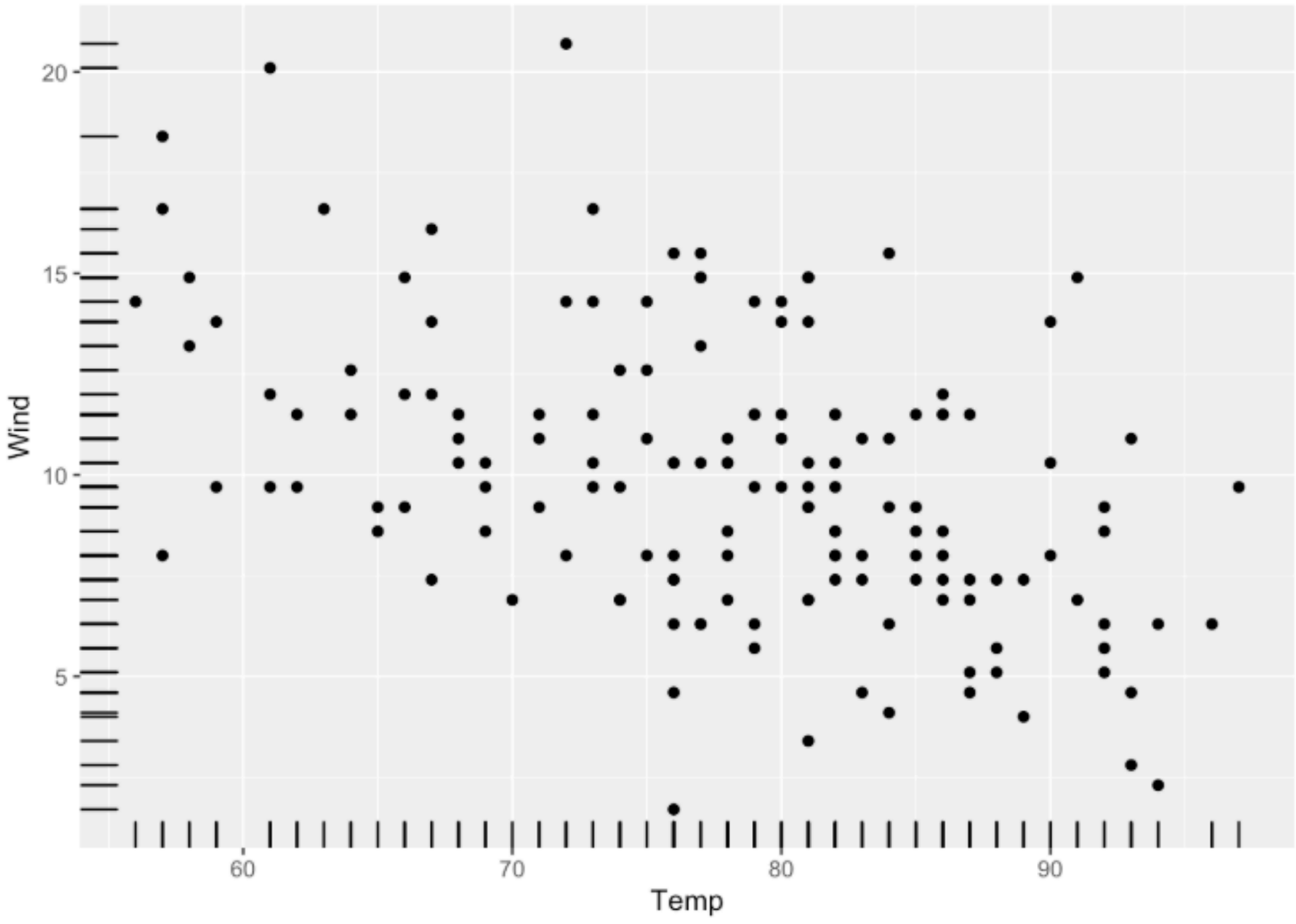
Trendlines over Points

```
ggplot( airquality, aes( x = Temp, y = Wind) ) + geom_point() + geom_smooth()
```



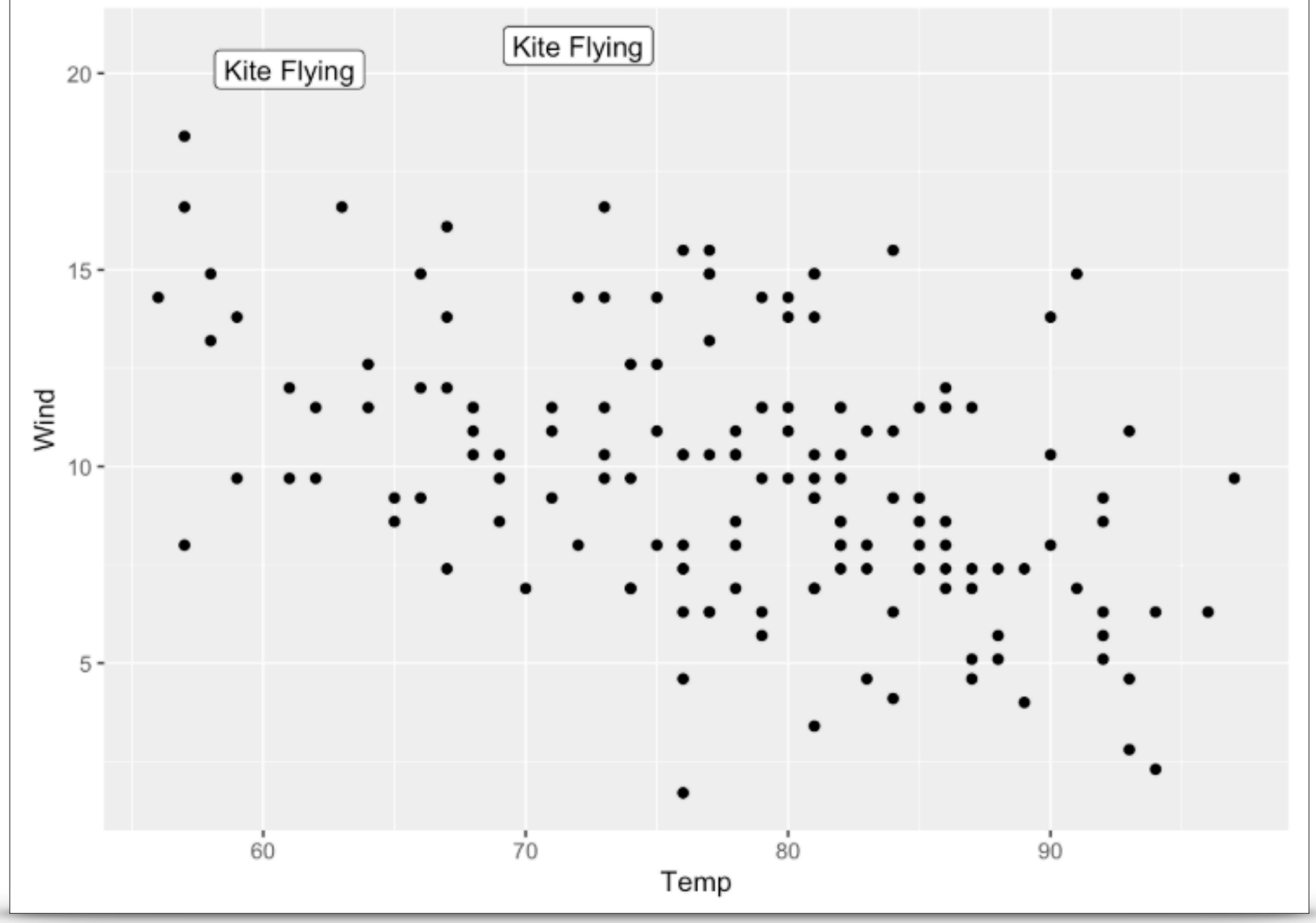
A Rug...

```
ggplot( airquality, aes( x = Temp, y = Wind) ) + geom_point() + geom_rug()
```



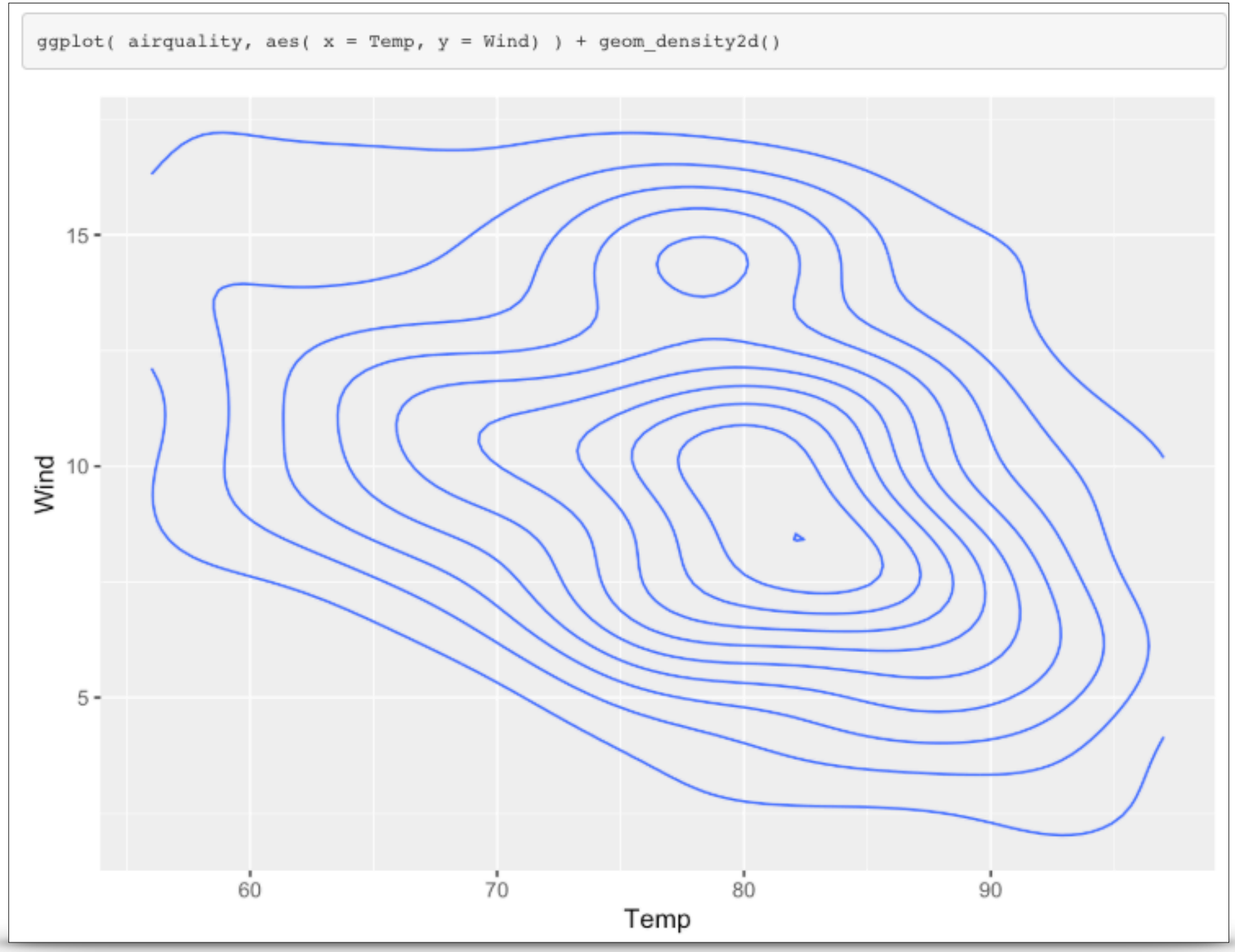
Adding Labels

```
ggplot( airquality, aes( x = Temp, y = Wind) ) +
  geom_point()  +
  geom_label( aes(label=Label), data = kite.days, position="nudge")
```



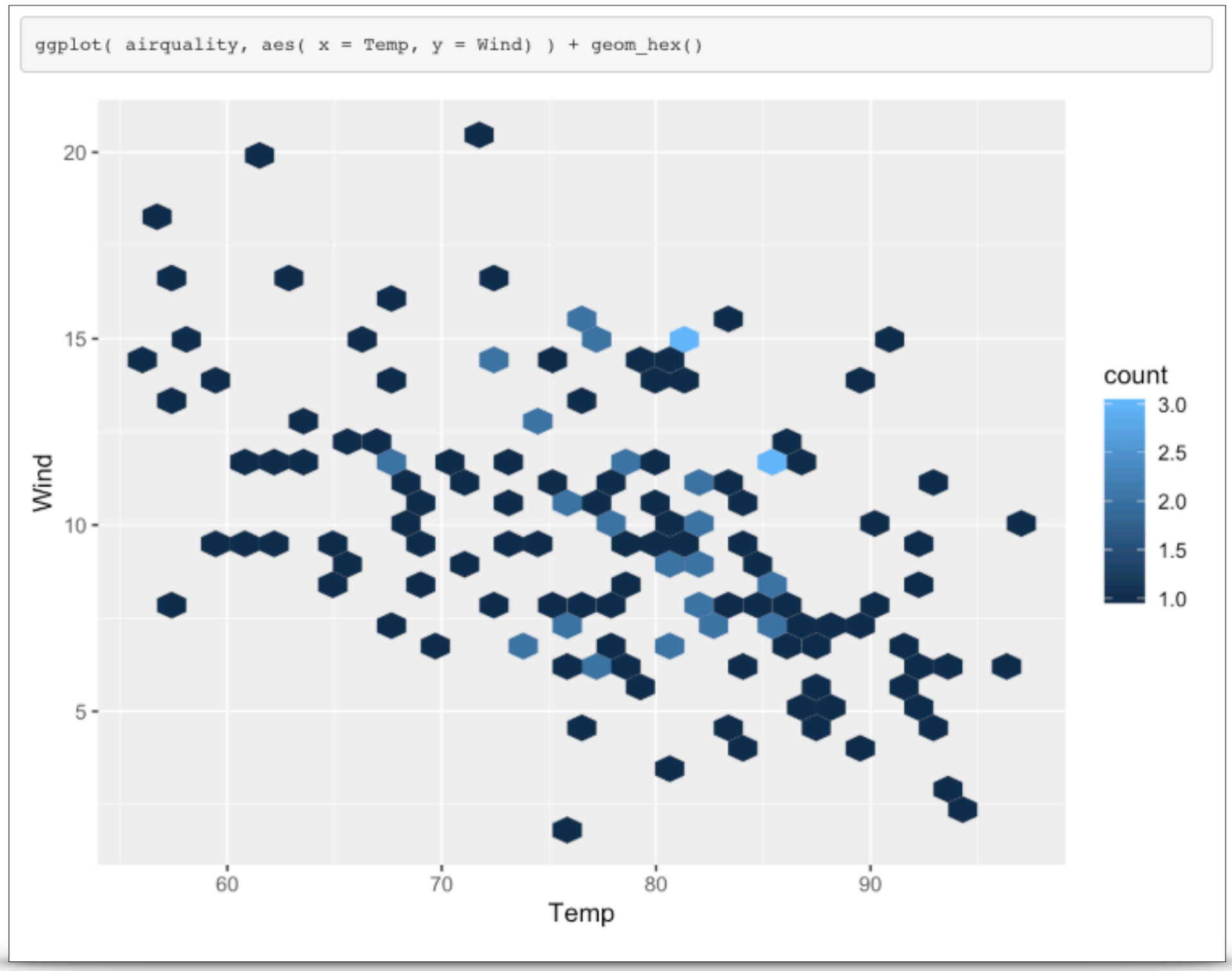
Contour Estimates

```
ggplot( airquality, aes( x = Temp, y = Wind) ) + geom_density2d()
```



Hex Plots

```
ggplot( airquality, aes( x = Temp, y = Wind) ) + geom_hex()
```

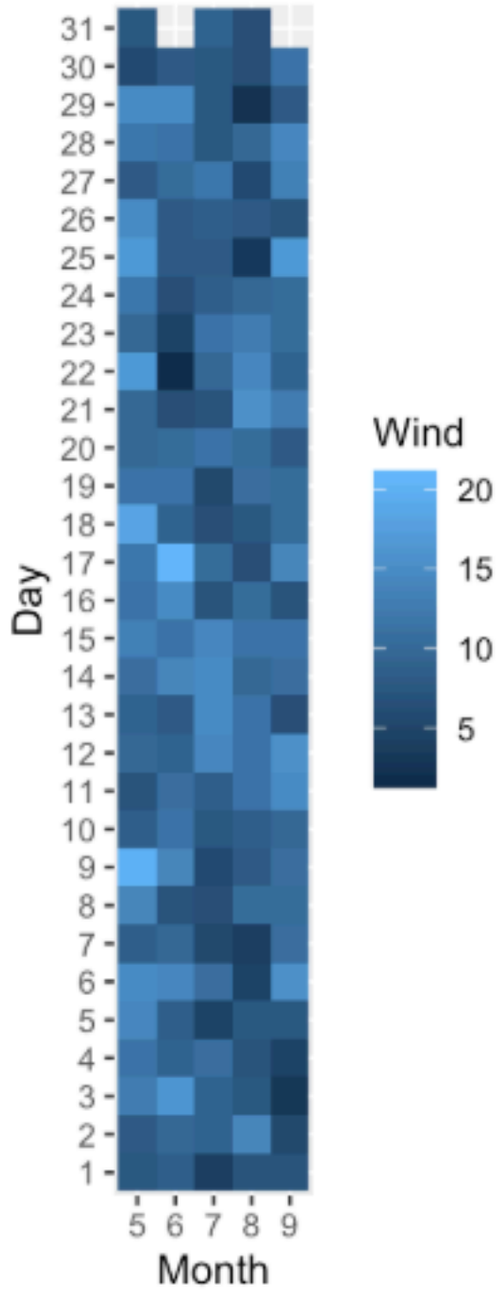


```
ggplot( data, aes(x=,y=)) + geom_***() + scale_x_***() + theme_*()
```

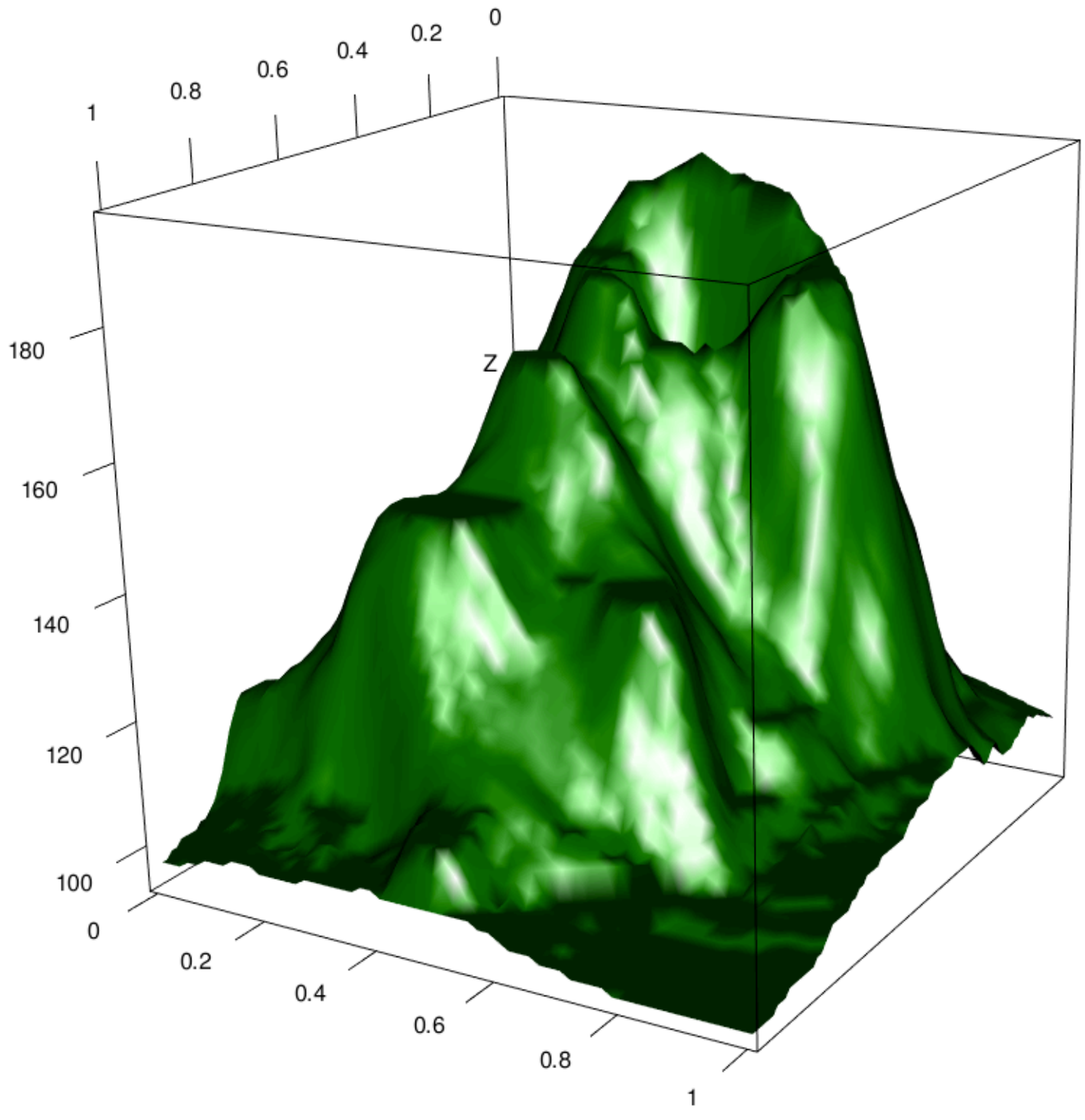
Geometries For 3 Variables

Tiles

```
# geom_tile() & geom_raster() make the same output
ggplot( airquality, aes( x=Month, y=Day, fill = Wind) ) + geom_tile() + coord_equal()
```



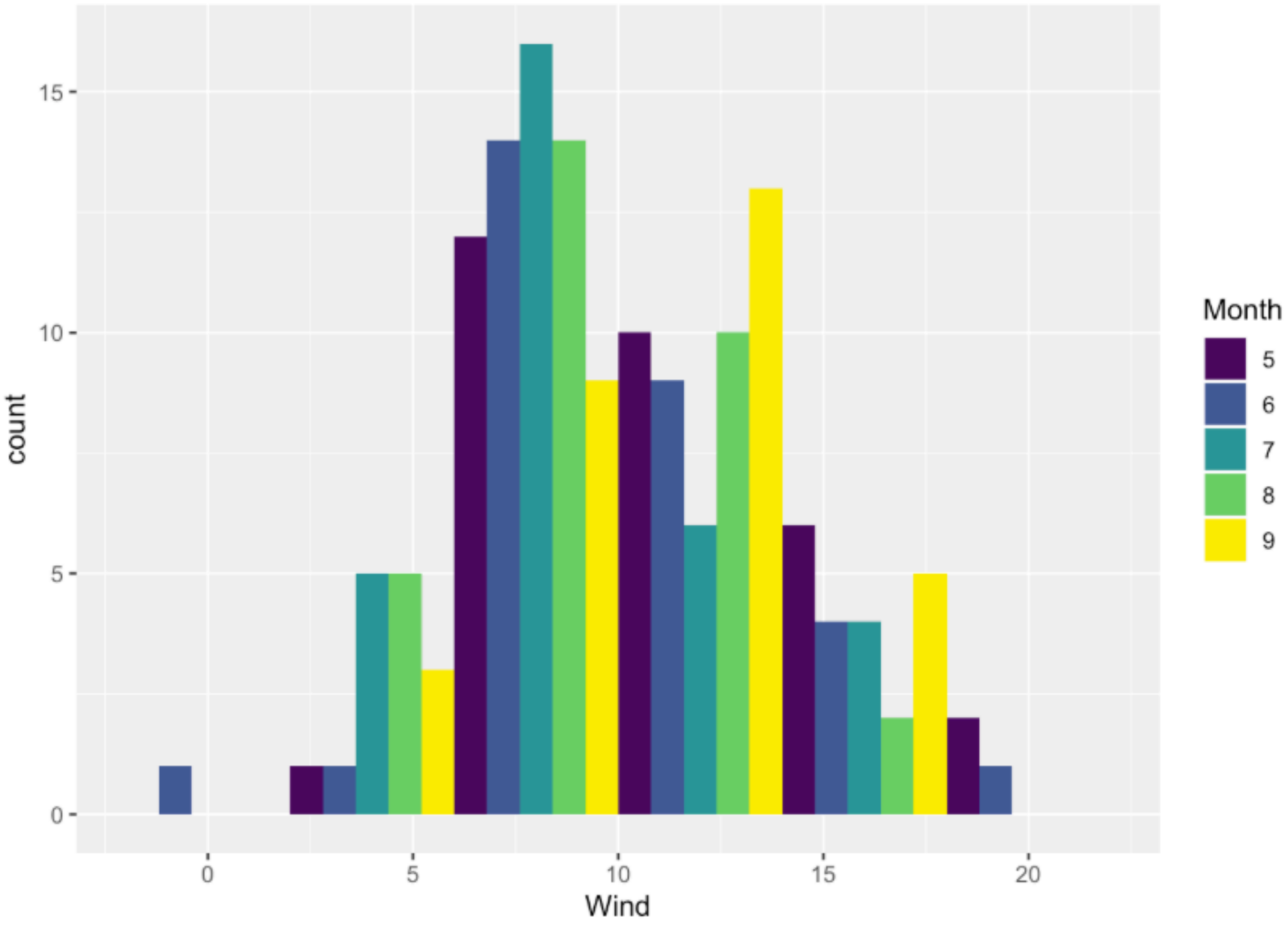
3D - OpenGL



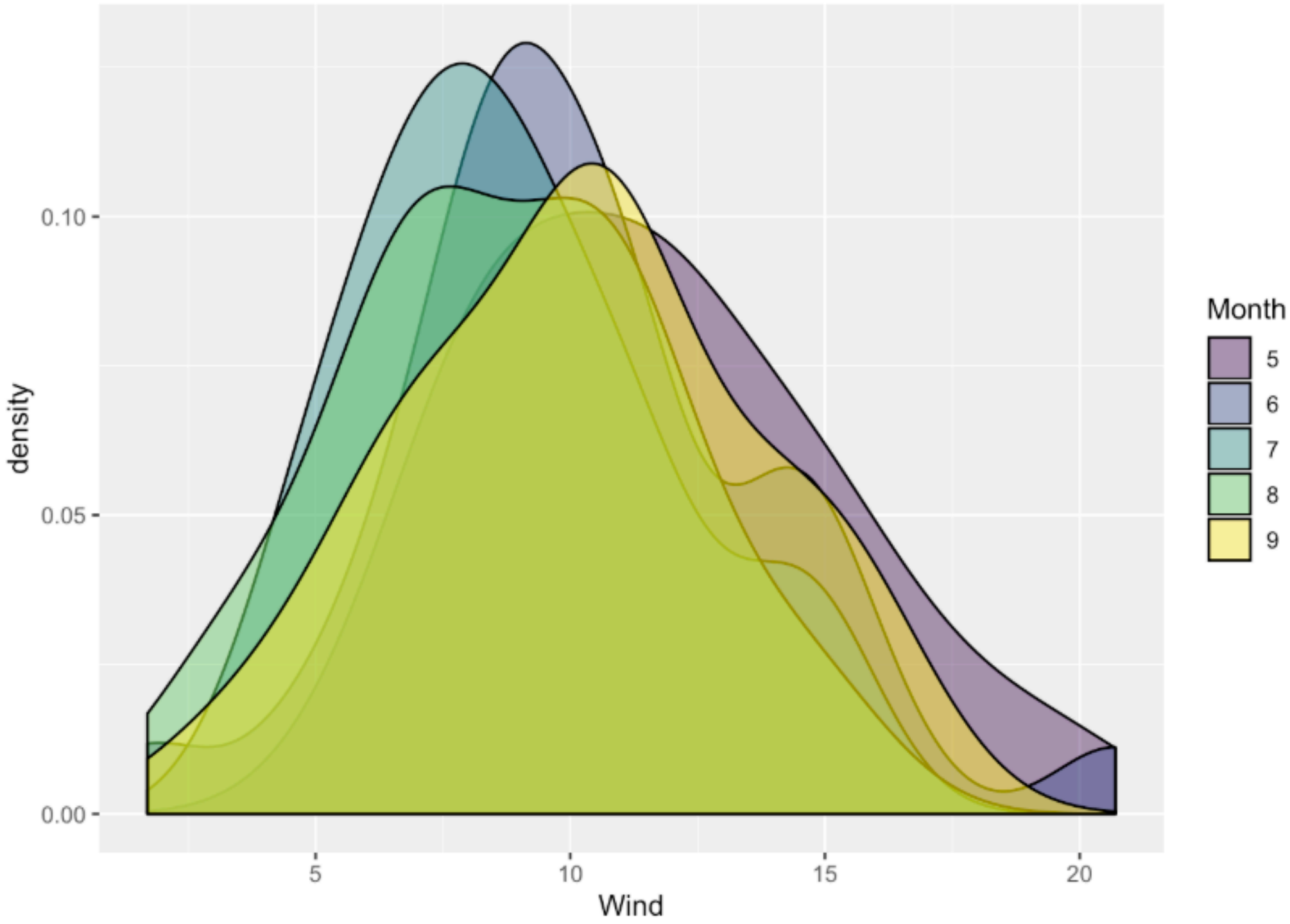
```
ggplot( data, aes(x=,y=)) + geom_***() + scale_x_***() + theme_*()
```

Transformations

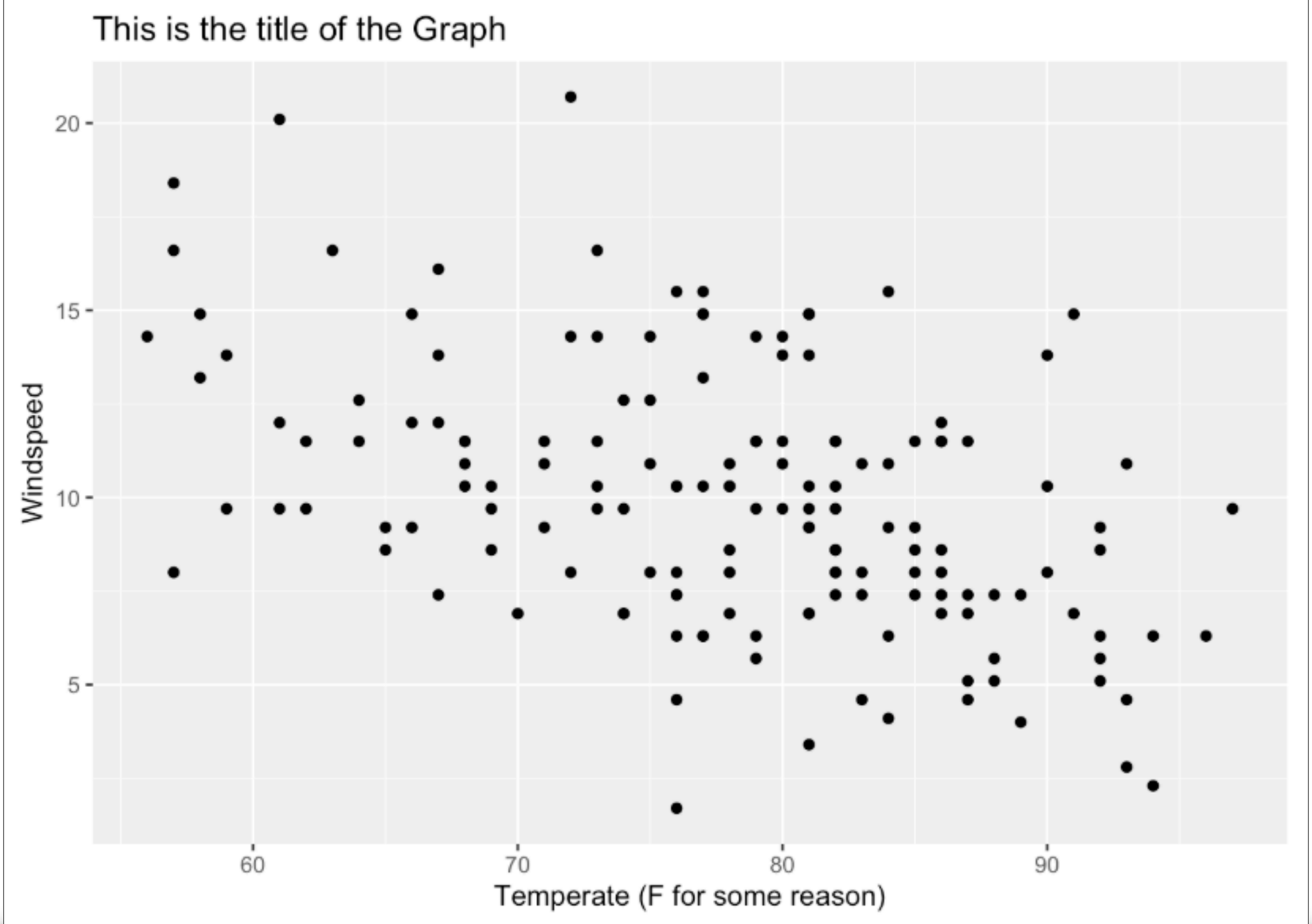
```
ggplot( airquality, aes(Wind, fill=Month) ) +
  geom_histogram(position="dodge", binwidth=4)
```



```
ggplot( airquality, aes(Wind, fill=Month) ) +
  geom_density( alpha=0.5 )
```

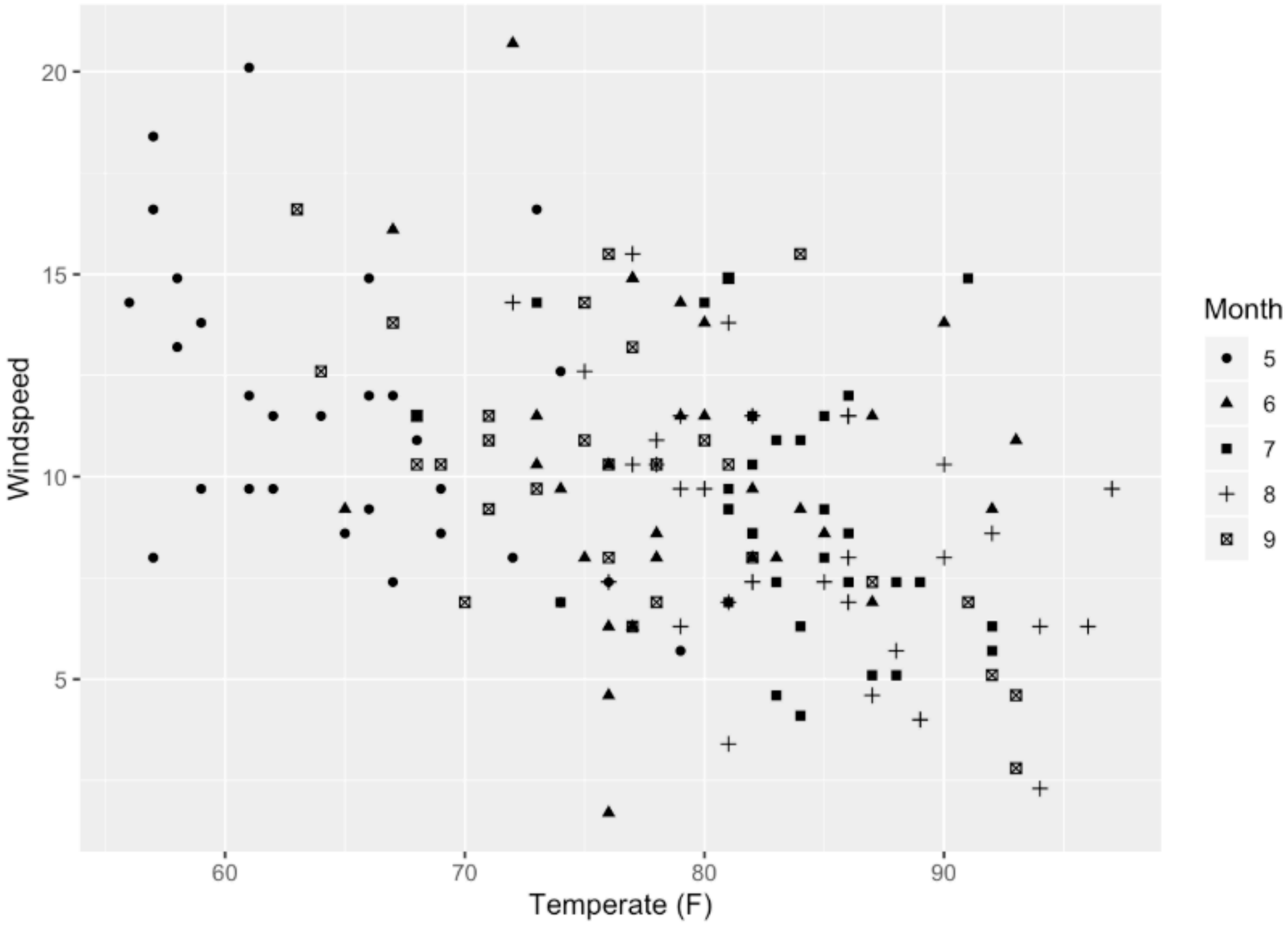


```
ggplot( airquality, aes(Temp, Wind)) + geom_point() +
  ylab("Windspeed") +
  xlab("Temperate (F for some reason)") +
  ggtitle("This is the title of the Graph")
```

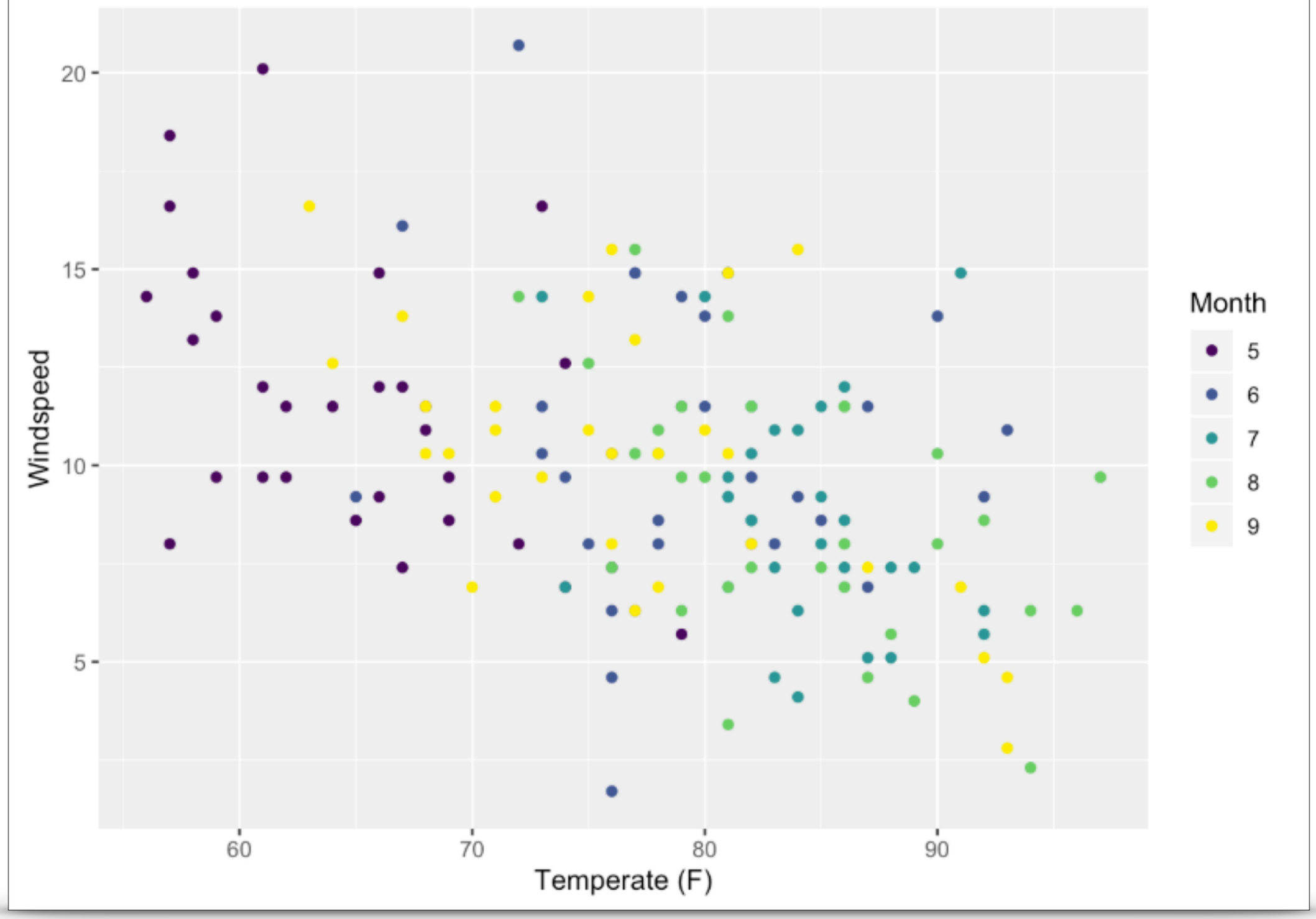


```
ggplot( airquality, aes(Temp, Wind, shape=Month)) + geom_point() +
  ylab("Windspeed") +
  xlab("Temperate (F)")
```

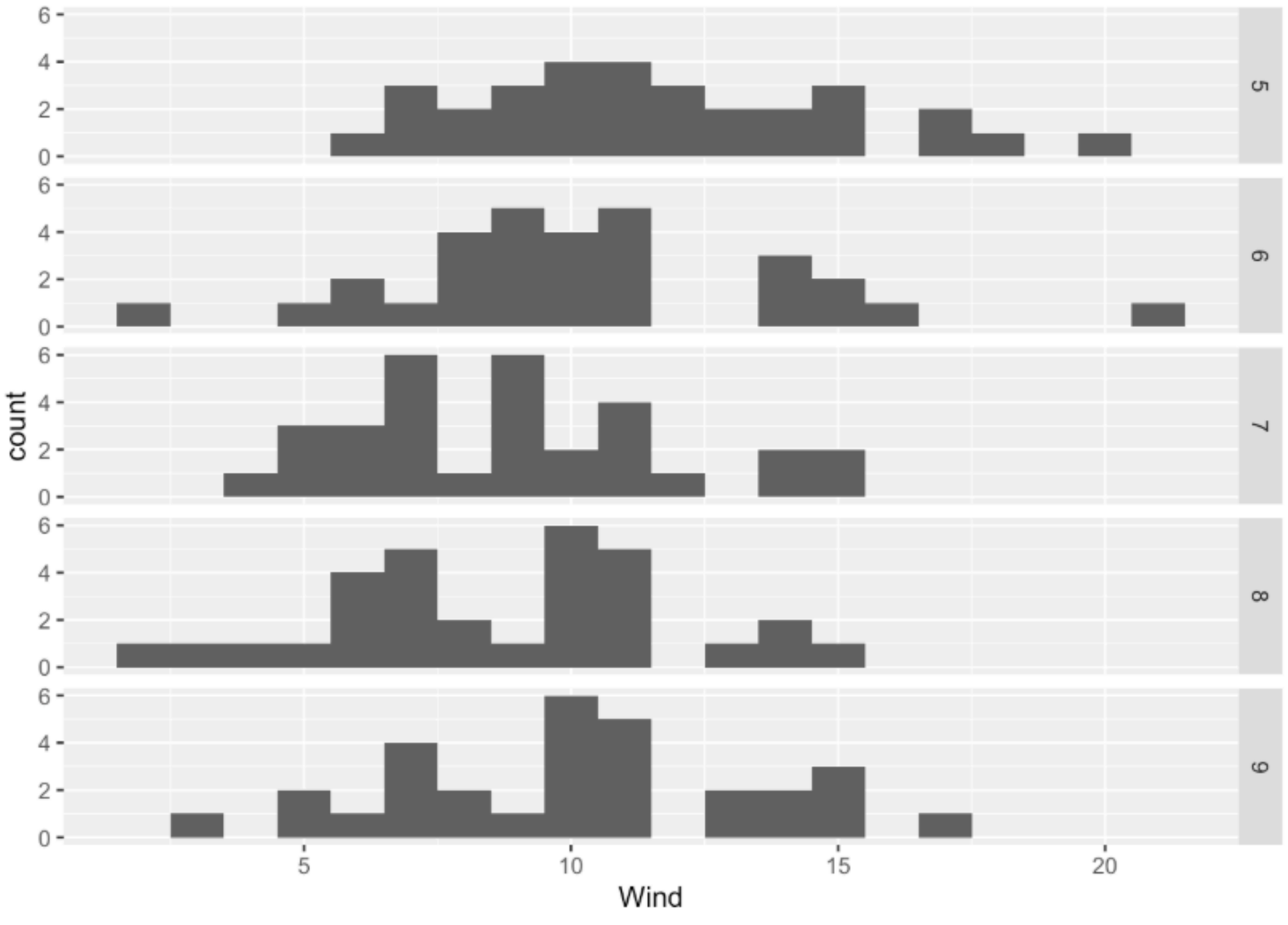
```
## Warning: Using shapes for an ordinal variable is not advised
```



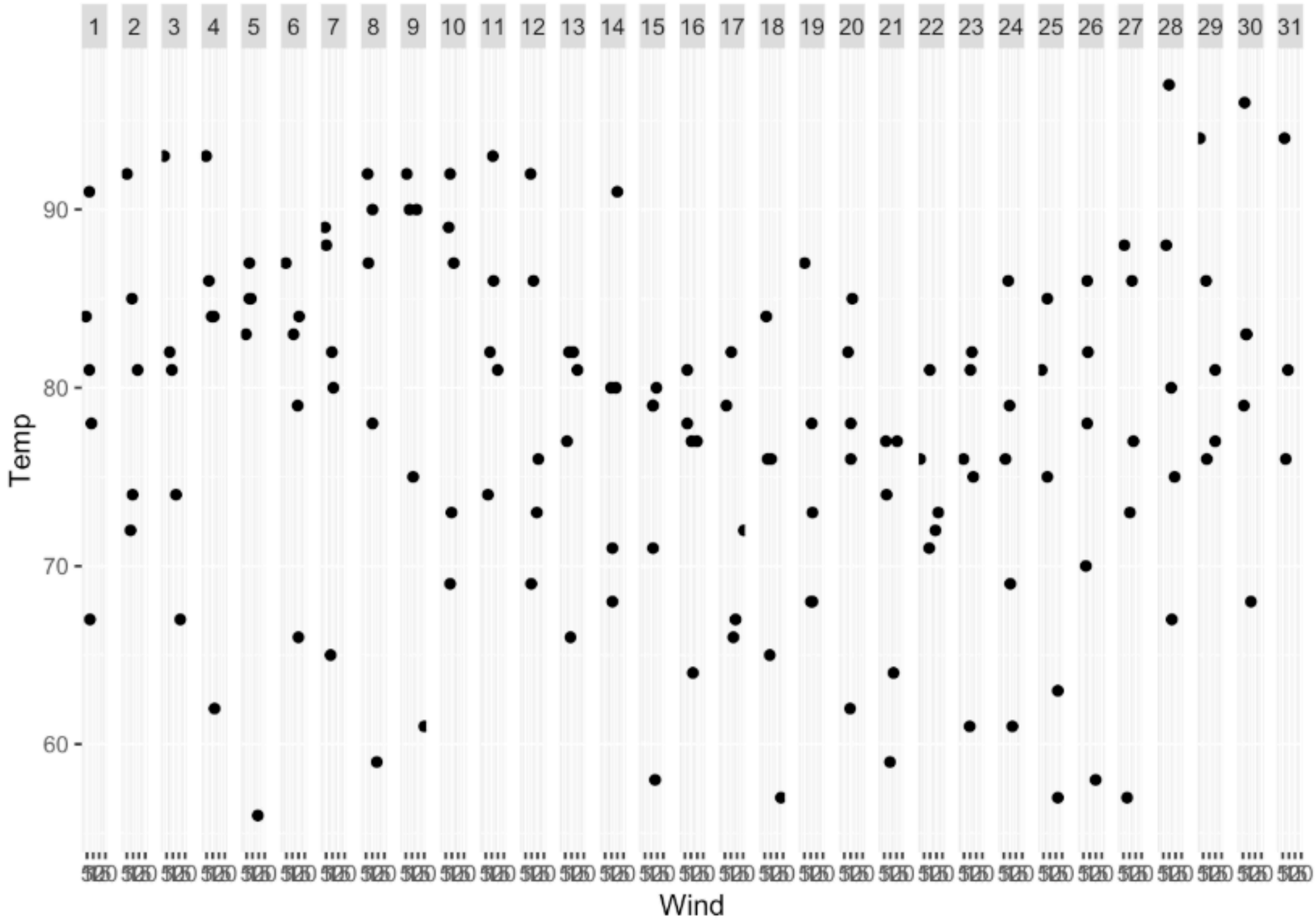
```
ggplot( airquality, aes(Temp, Wind, color=Month)) + geom_point() +
  ylab("Windspeed") +
  xlab("Temperate (F)")
```



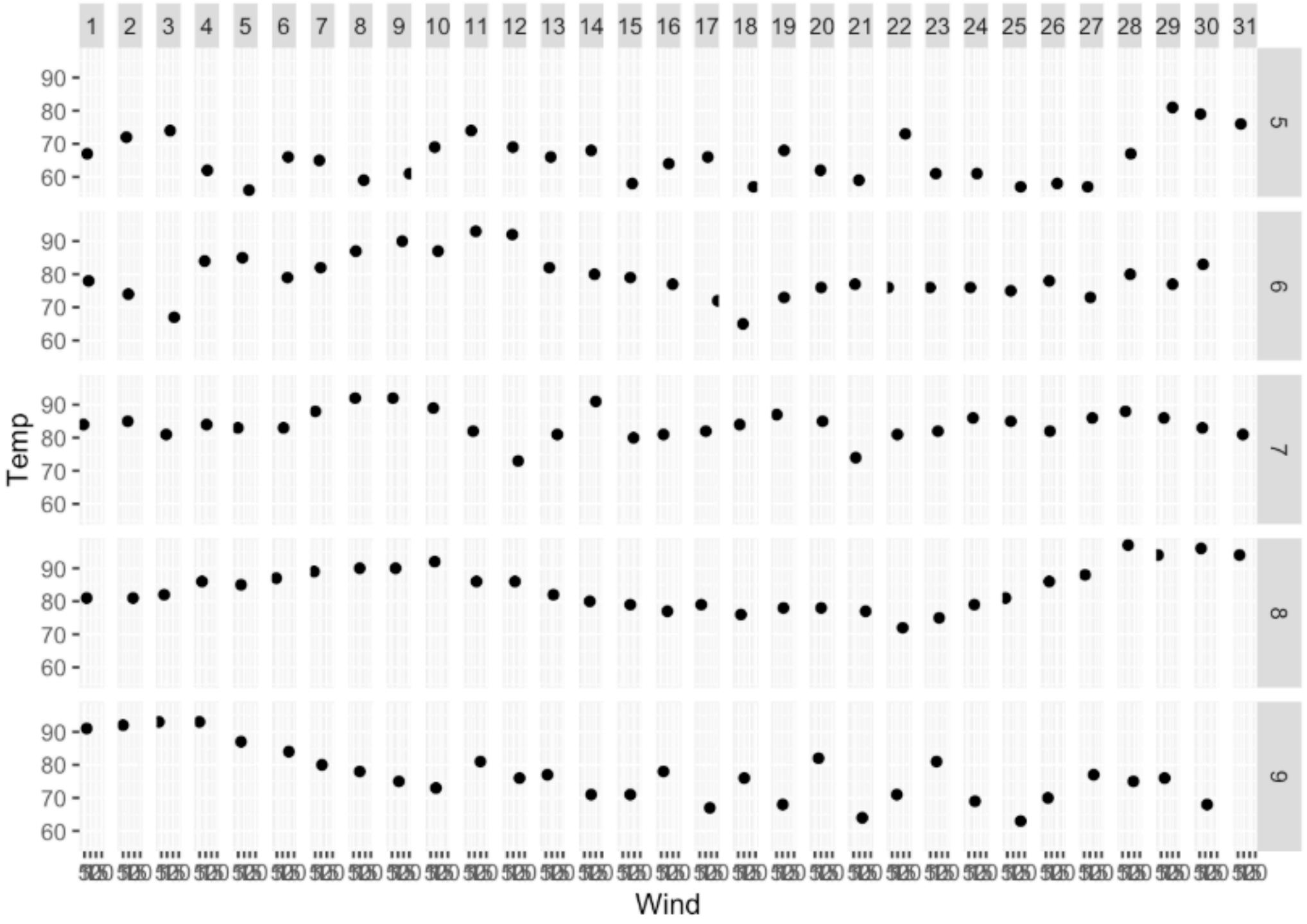
```
ggplot( airquality, aes(Wind)) +
  geom_histogram( binwidth=1.0) +
  facet_grid( Month ~ .)
```



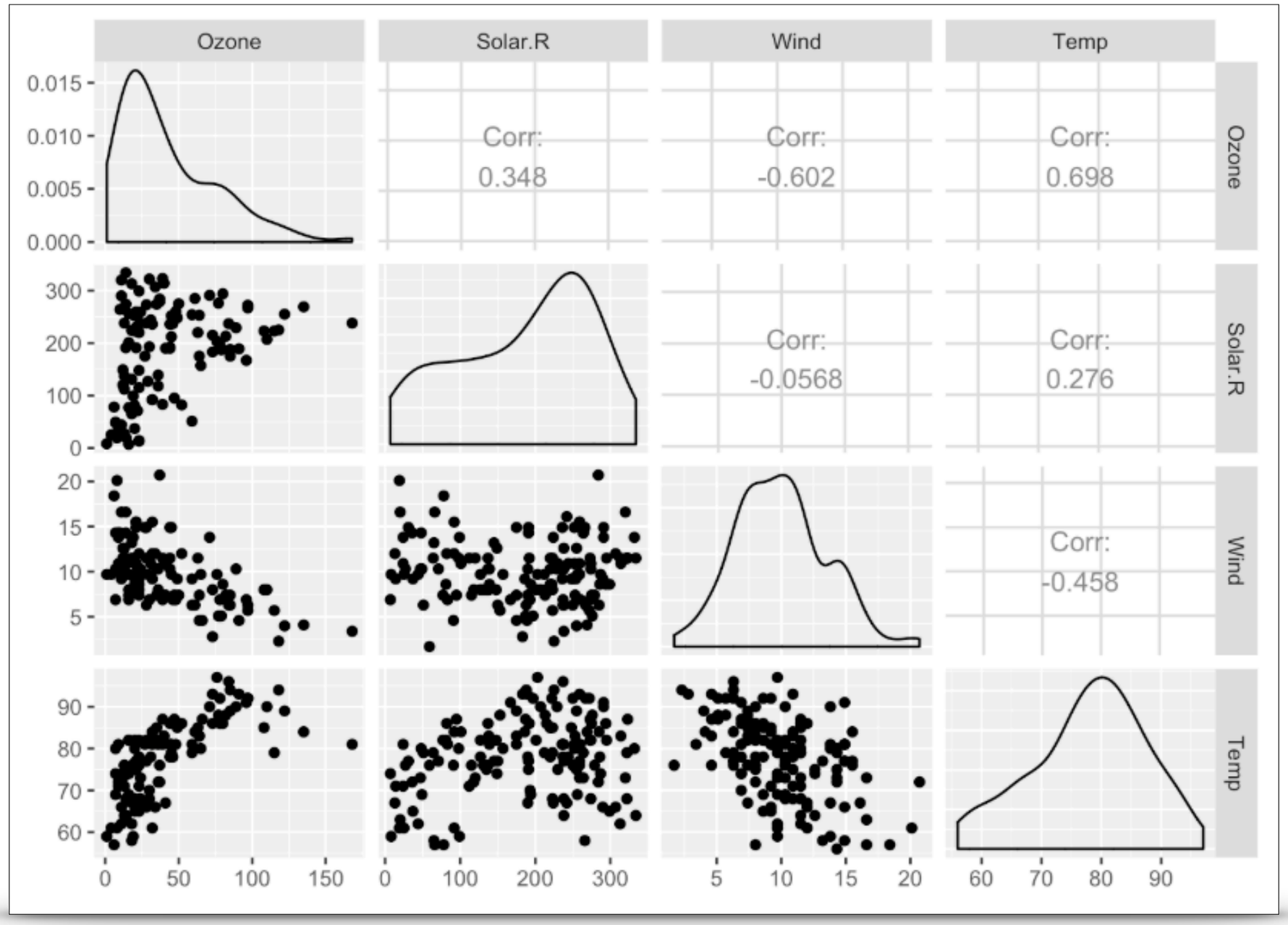
```
ggplot( airquality, aes(Wind, Temp)) +
  geom_point() +
  facet_grid( . ~ Day)
```



```
ggplot( airquality, aes(Wind,Temp)) +
  geom_point() +
  facet_grid( Month ~ Day)
```



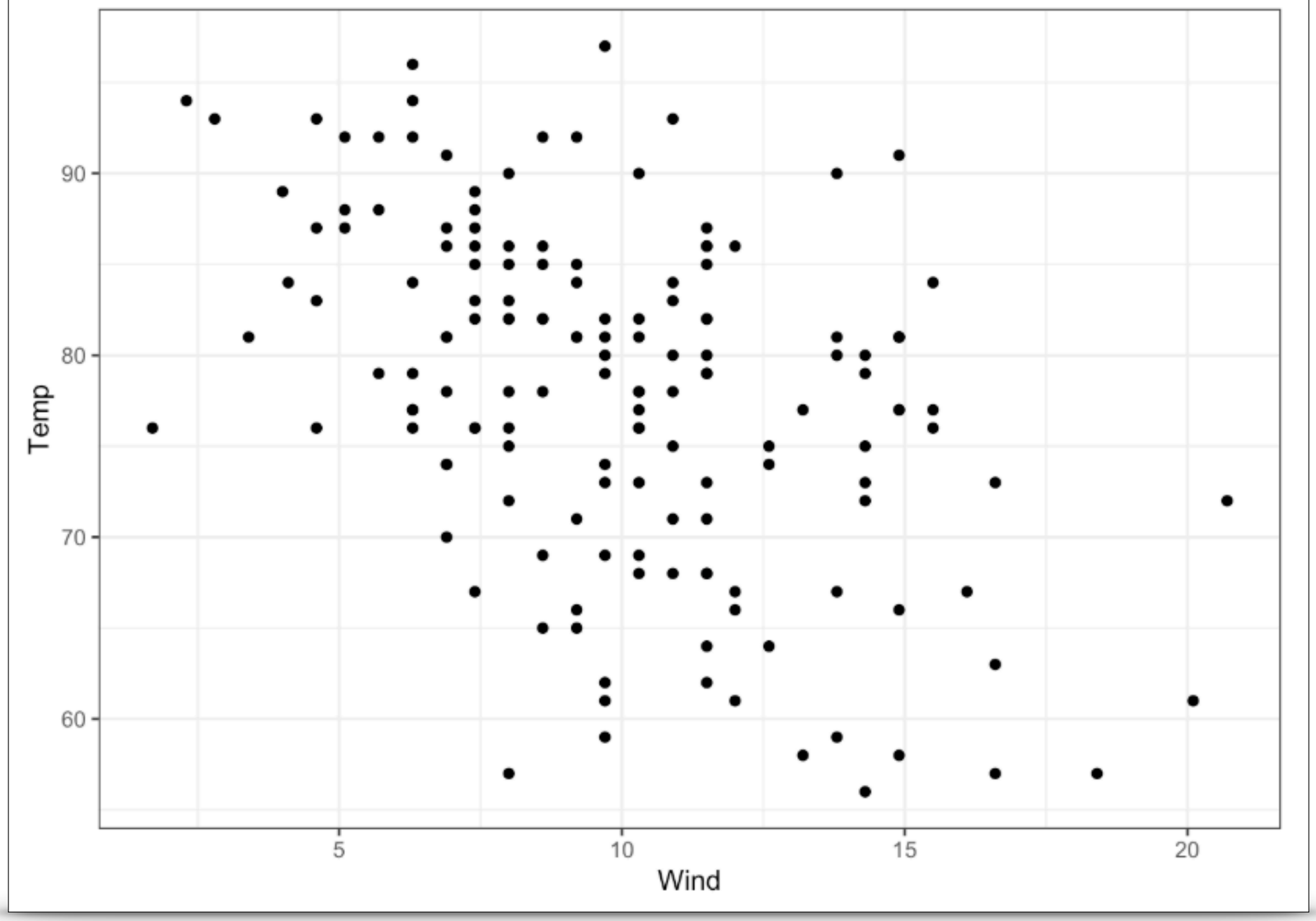
```
library(GGally)
airquality %>%
  select( Ozone, Solar.R, Wind, Temp ) %>%
  ggpairs()
```



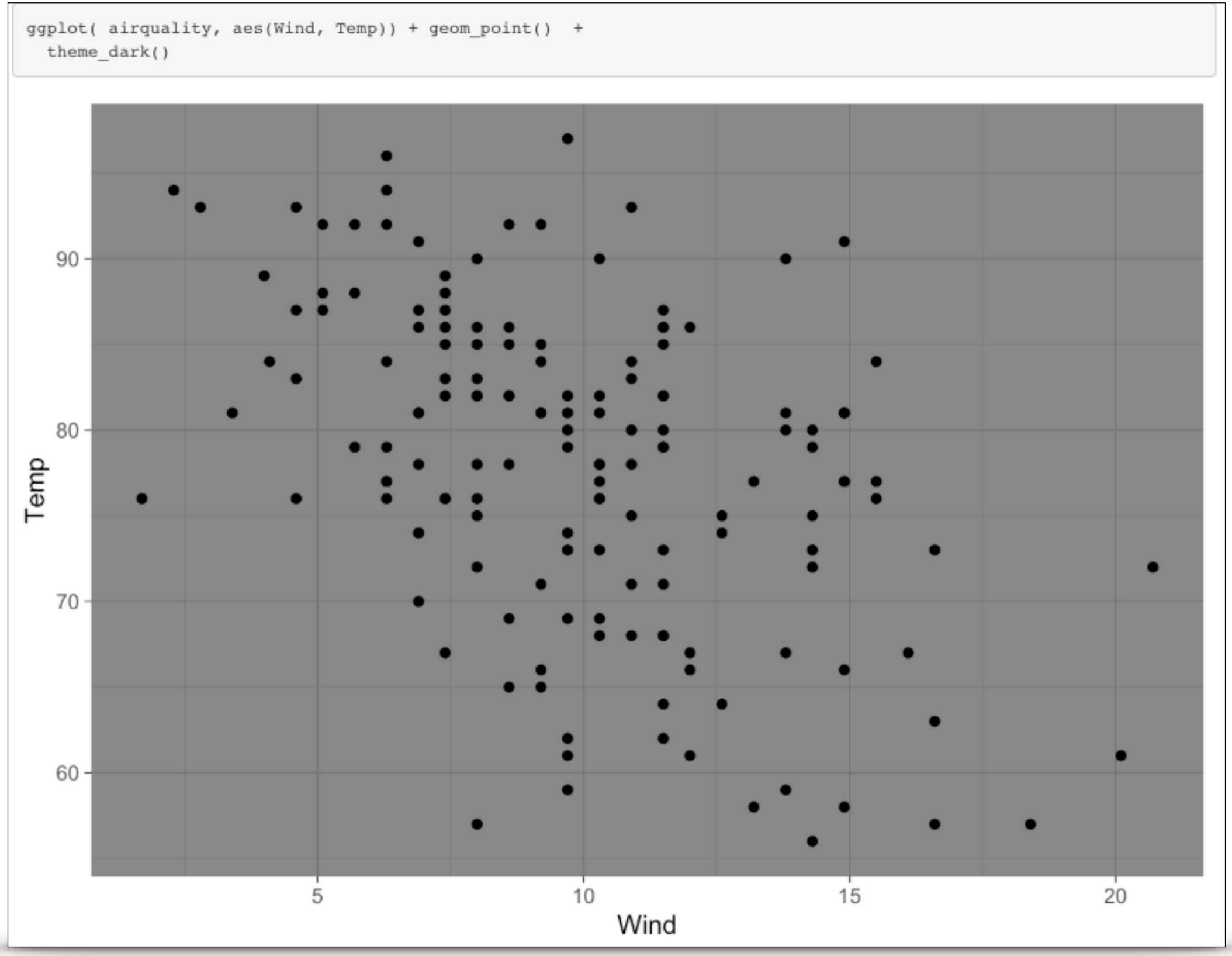
```
ggplot( data, aes(x=,y=)) + geom_***() + scale_x_***() + theme_*()
```

Themes

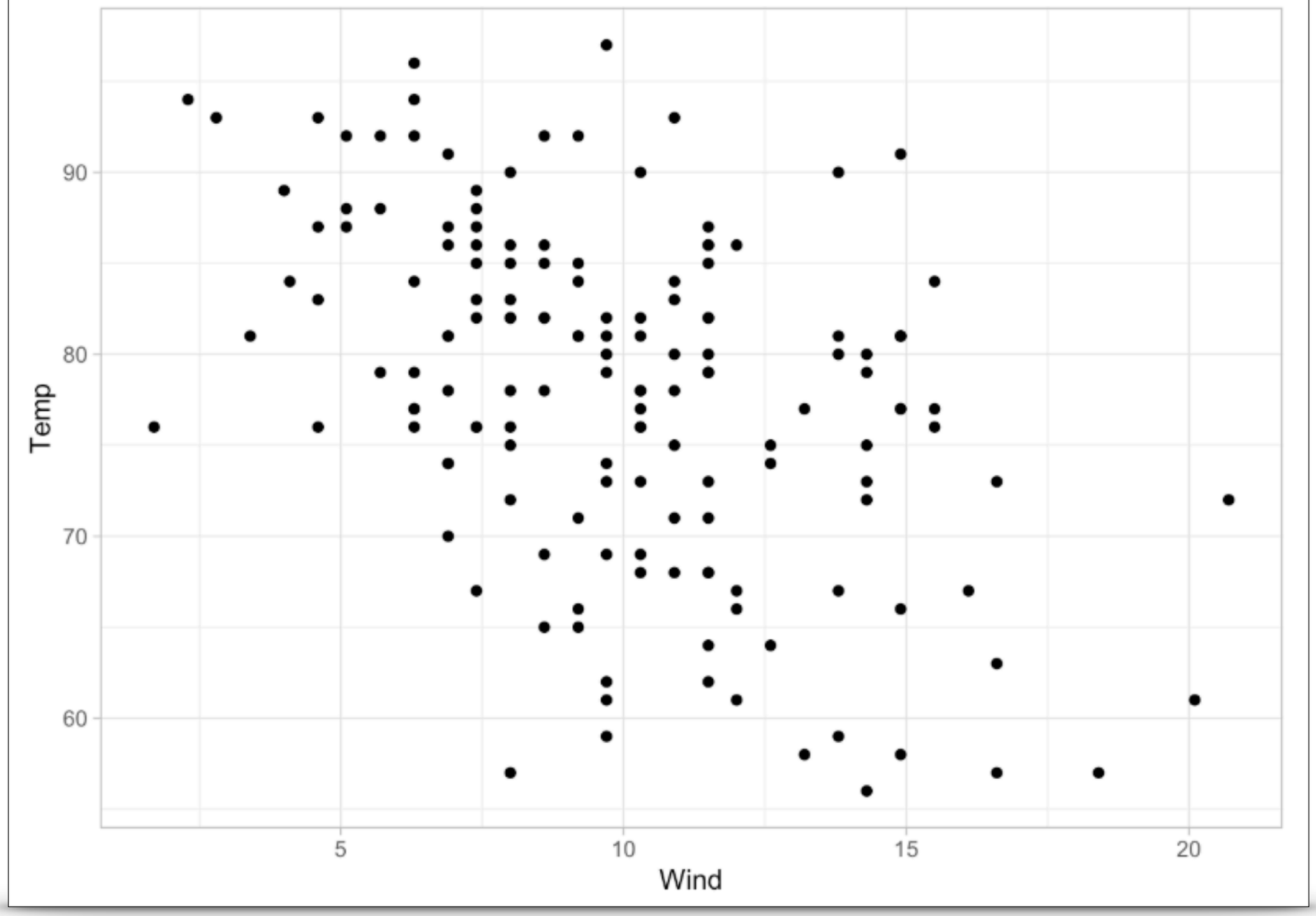
```
ggplot( airquality, aes(Wind, Temp)) + geom_point()  +
  theme_bw()
```



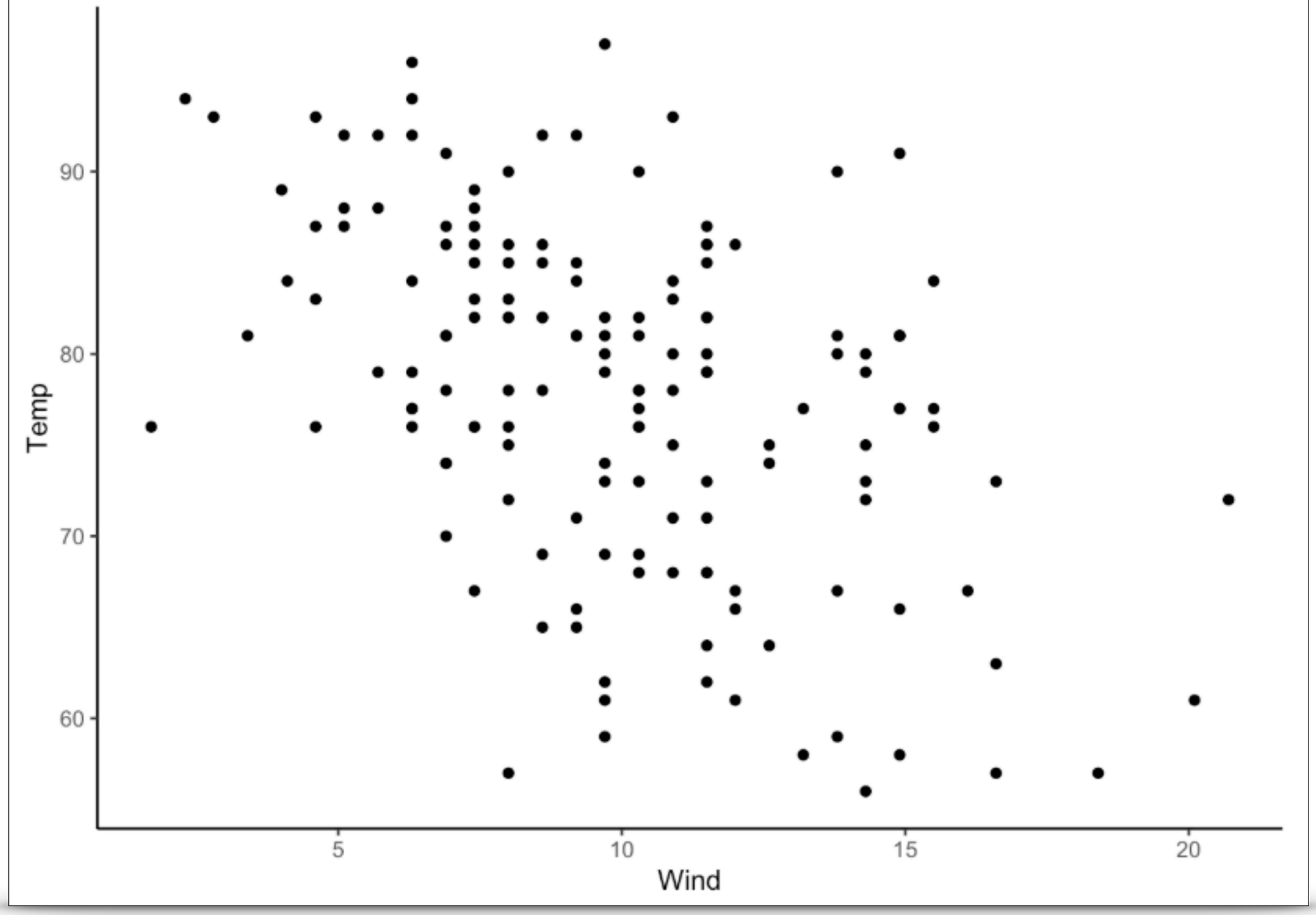
```
ggplot( airquality, aes(Wind, Temp)) + geom_point()  +
  theme_dark()
```



```
ggplot( airquality, aes(Wind, Temp)) + geom_point()  +
  theme_light()
```



```
ggplot( airquality, aes(Wind, Temp)) + geom_point()  +
  theme_classic()
```



```
ggplot( airquality, aes(Wind, Temp)) + geom_point()  +
  theme_void()
```