

Data Visualization

Representing Data With Colors

Prof. Roger D. Peng

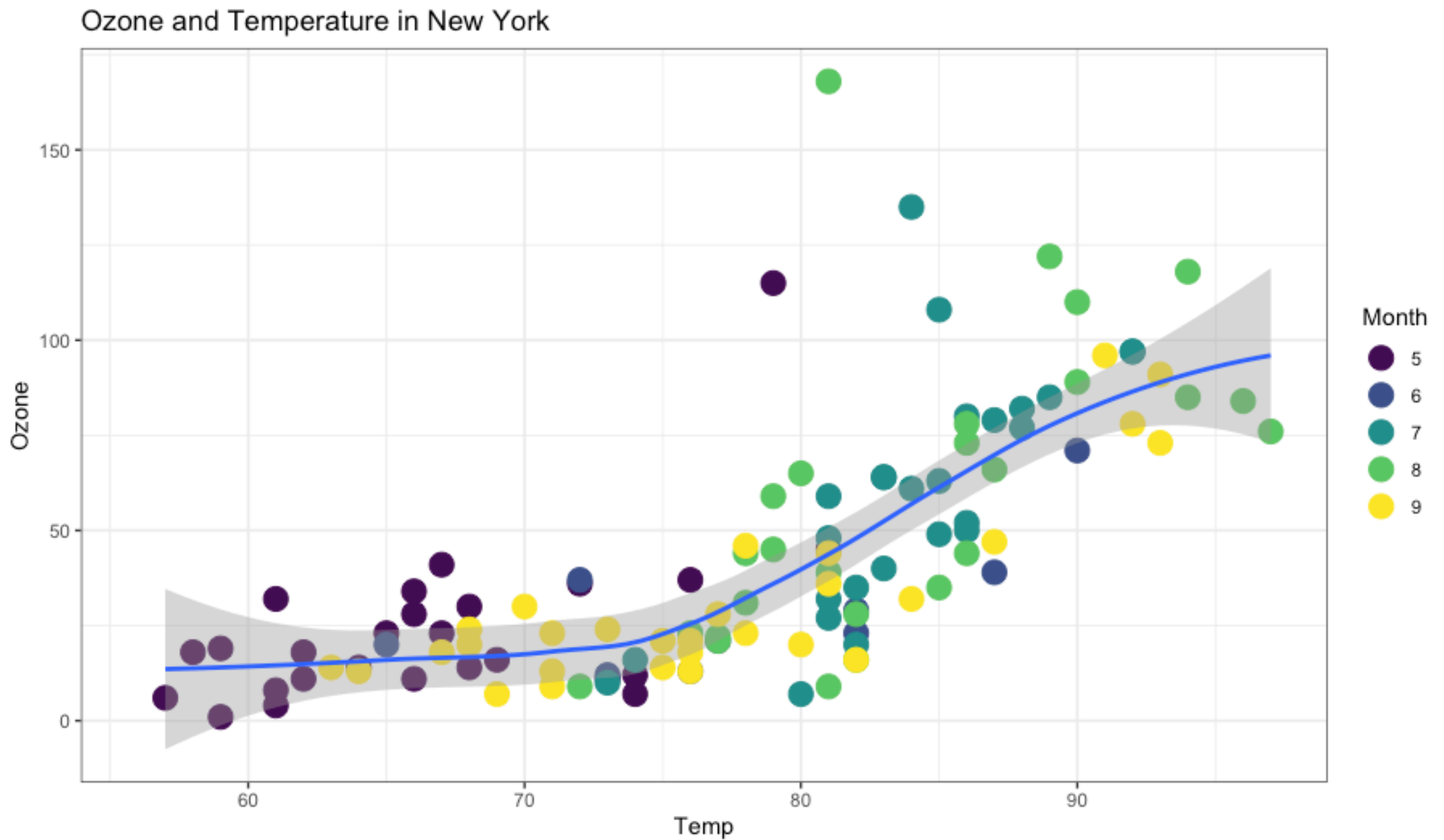
Department of Statistics and Data Sciences

University of Texas at Austin

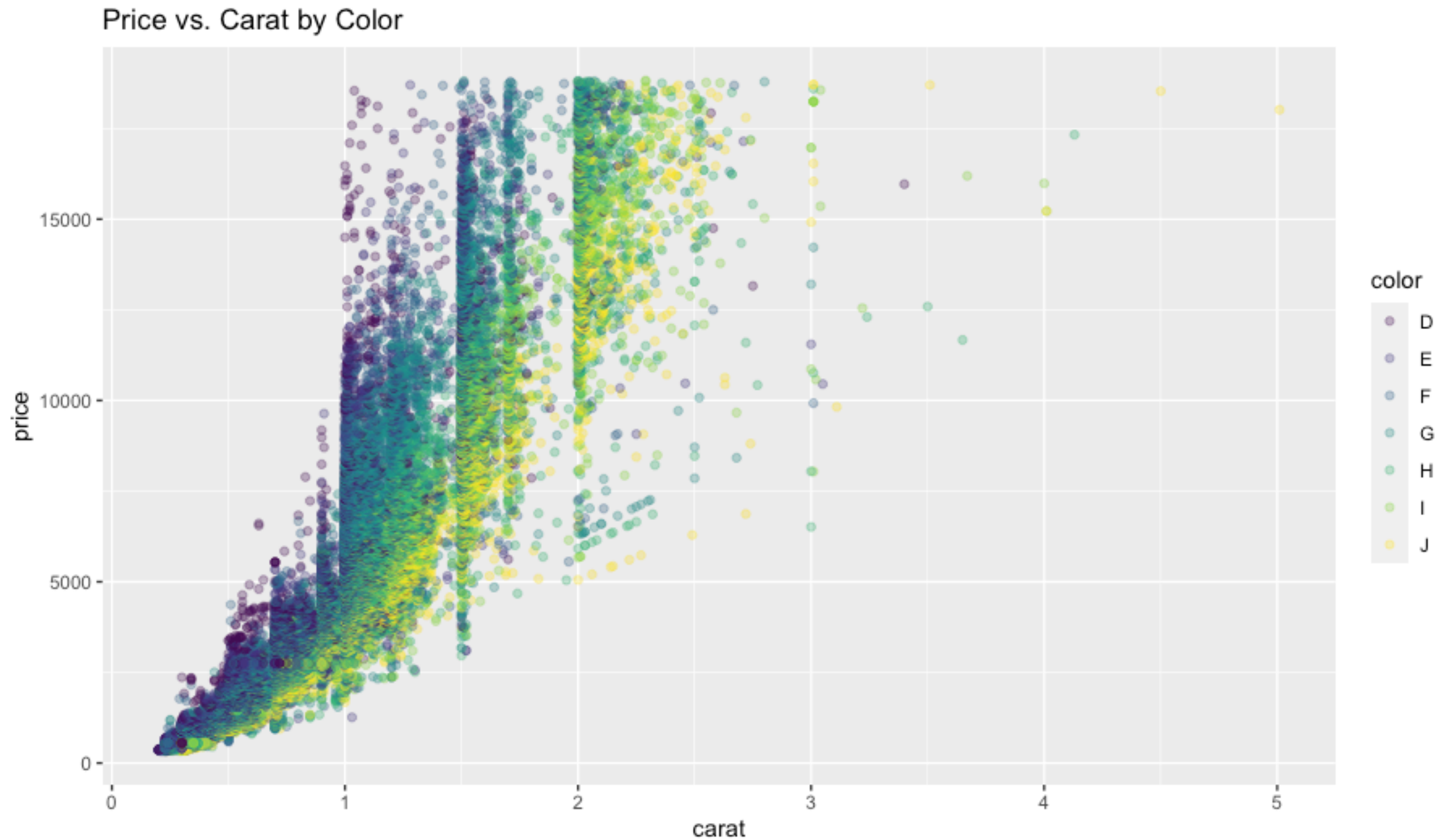
Why Color Is Important

- Employing different colors is a useful way to add another dimension of information to your plots and data graphics
- Color is a key tool for allowing us to "escape flatland"
- BUT: Certain palettes of color are more appropriate for certain types of data
- Some careful thinking about color can greatly enhance a data graphic

Escaping Flatland



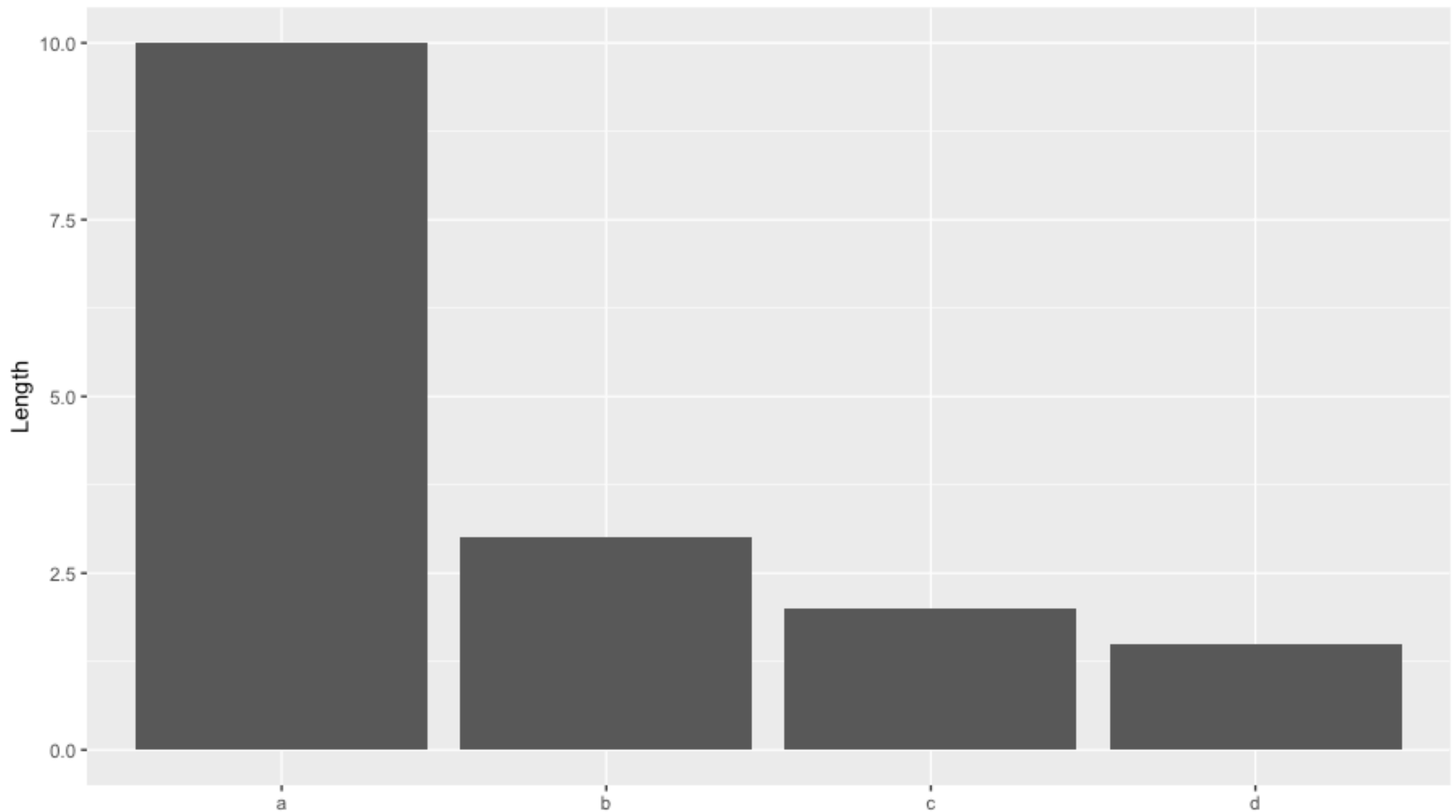
Escaping Flatland



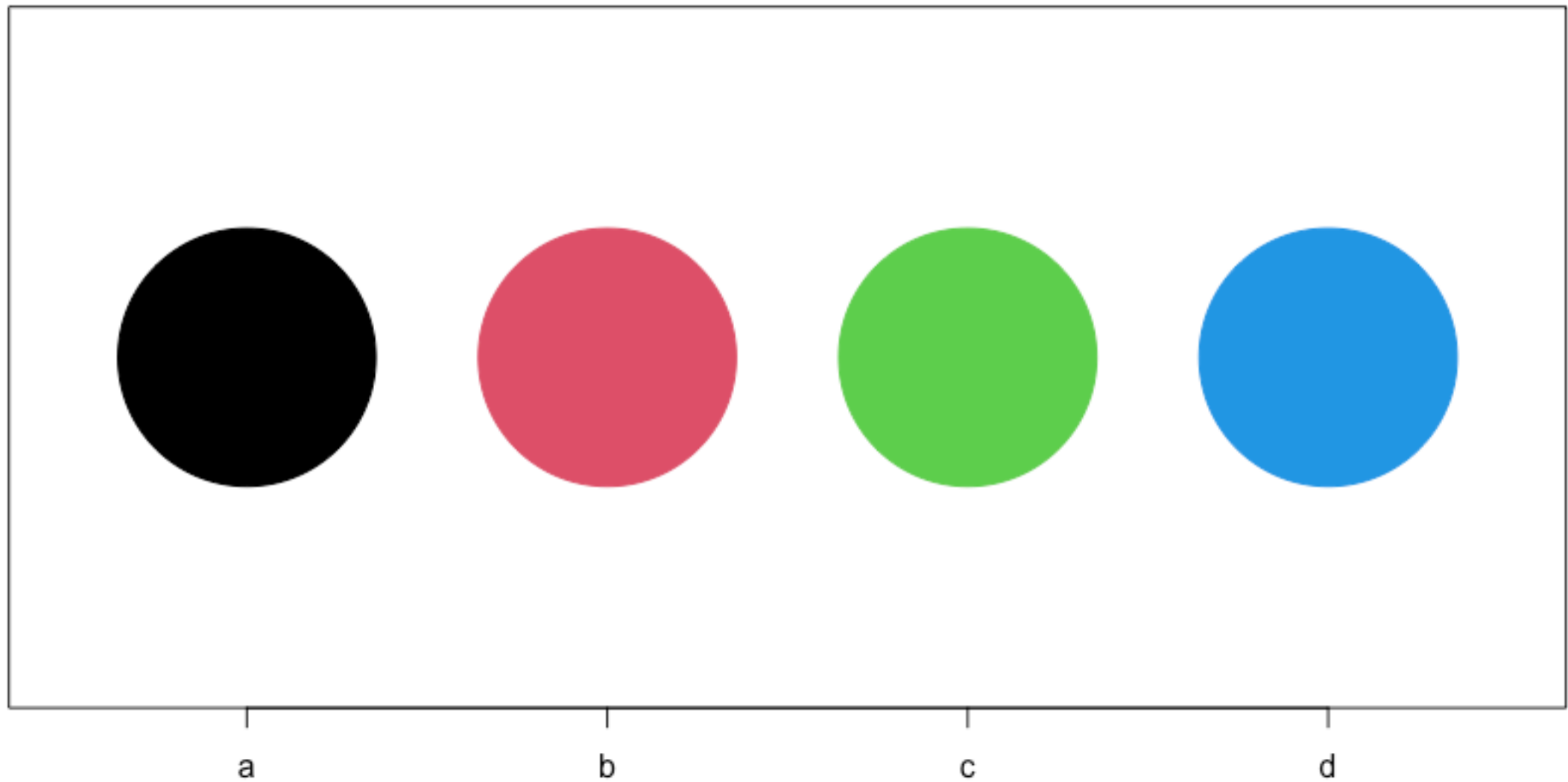
Data Graphics and Color

- Default colors for ggplot2 plots are okay but occasionally suboptimal
- There are functions in R and in external packages that are very handy for dealing with plotting colors
- **RColorBrewer** and the **viridis** packages provide interesting color palettes
- **scale_color_gradient()**, **scale_color_gradient2()**, and **scale_color_gradientn()** in ggplot2 allows you to manufacture your own color palettes

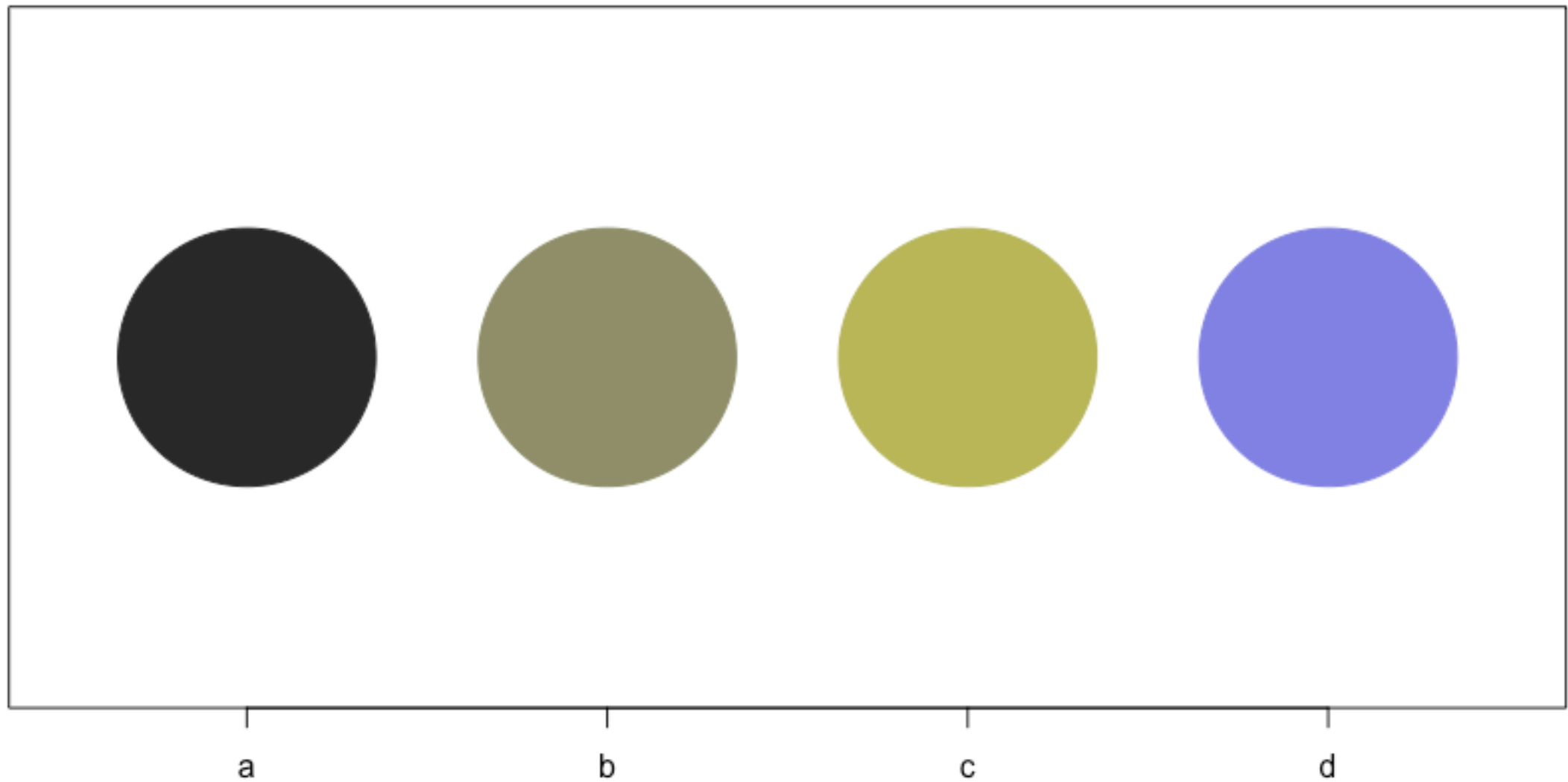
Perceiving Differences



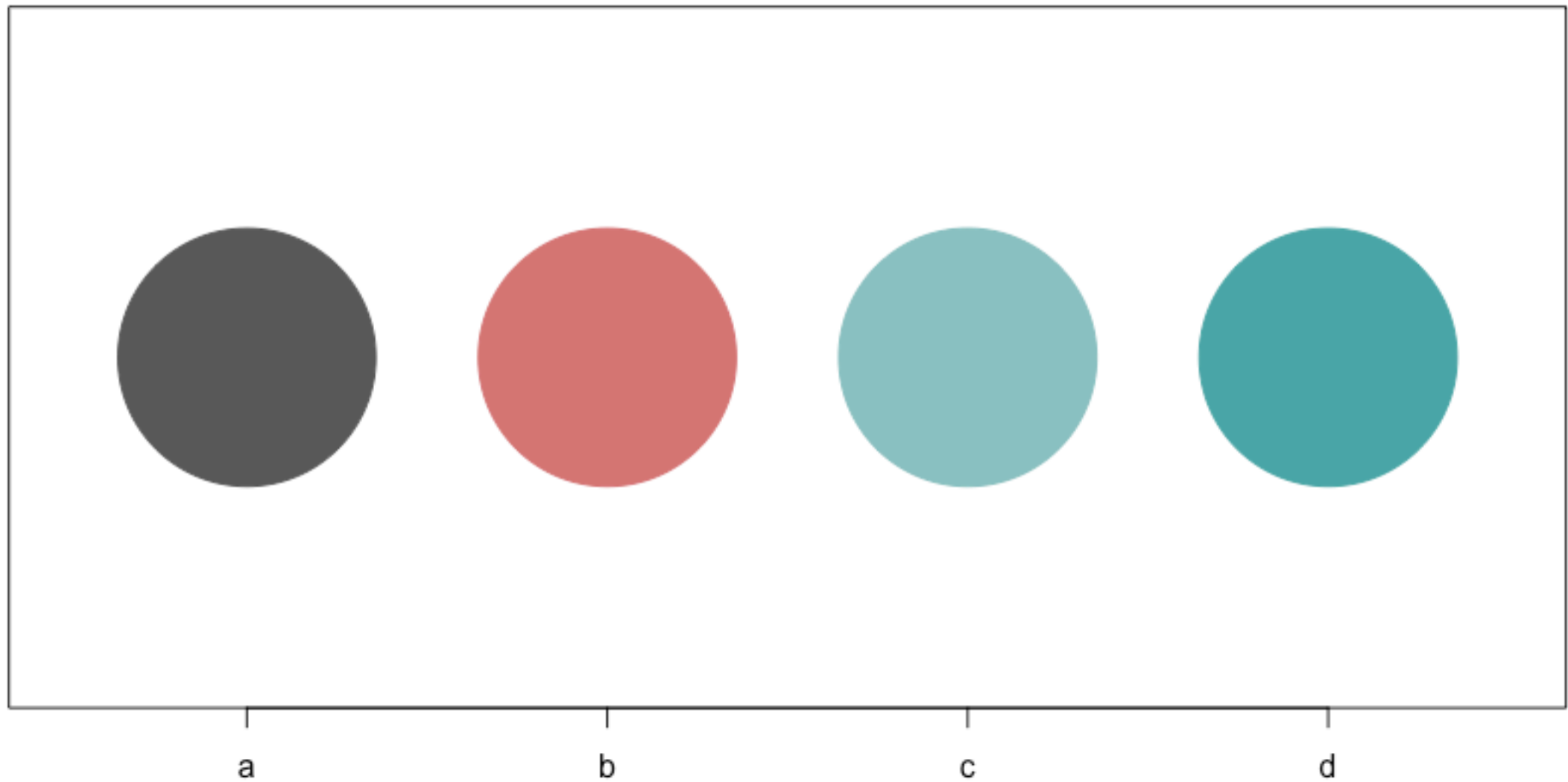
Perceiving Color Differences



Perceiving Color Differences



Perceiving Color Differences



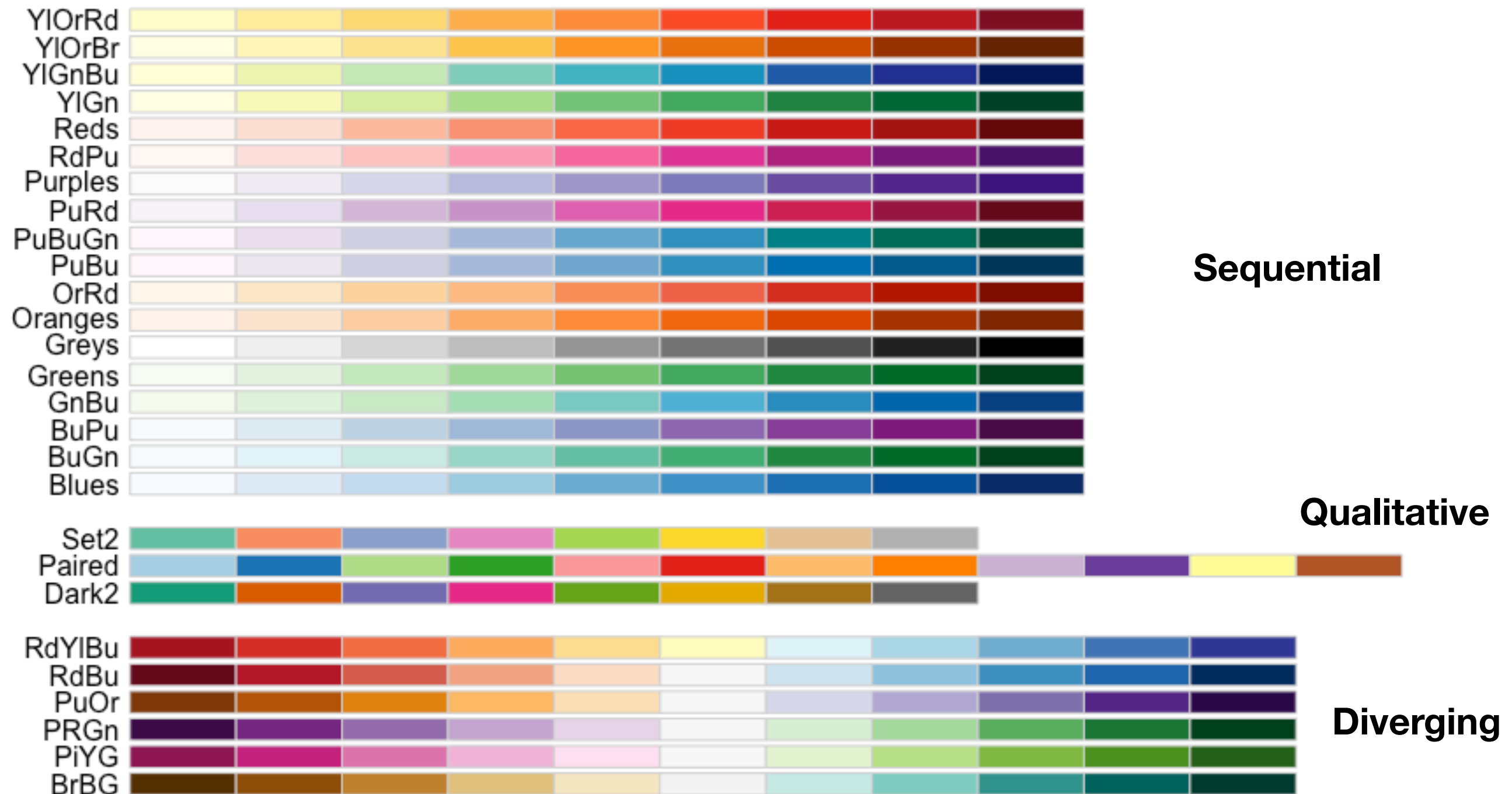
Representing Contrasts or Differences With Color

- Color can be usefully employed to represent differences in a separate dimension.
- Different types of data require different types of color scales:
 - **Sequential:** data are ordered from low to high (e.g. 0, 1, 2, ...)
 - **Diverging:** data increase outward from a mid-point (e.g. above/below average)
 - **Categorical:** data fall into a finite number of unordered categories (e.g. yes / no / maybe)
- Color scales should maximize range so that differences are easy to spot
- Perceptual uniformity, so that similar colors represent similar numbers, and dissimilar colors represent dissimilar numbers

RColorBrewer Package

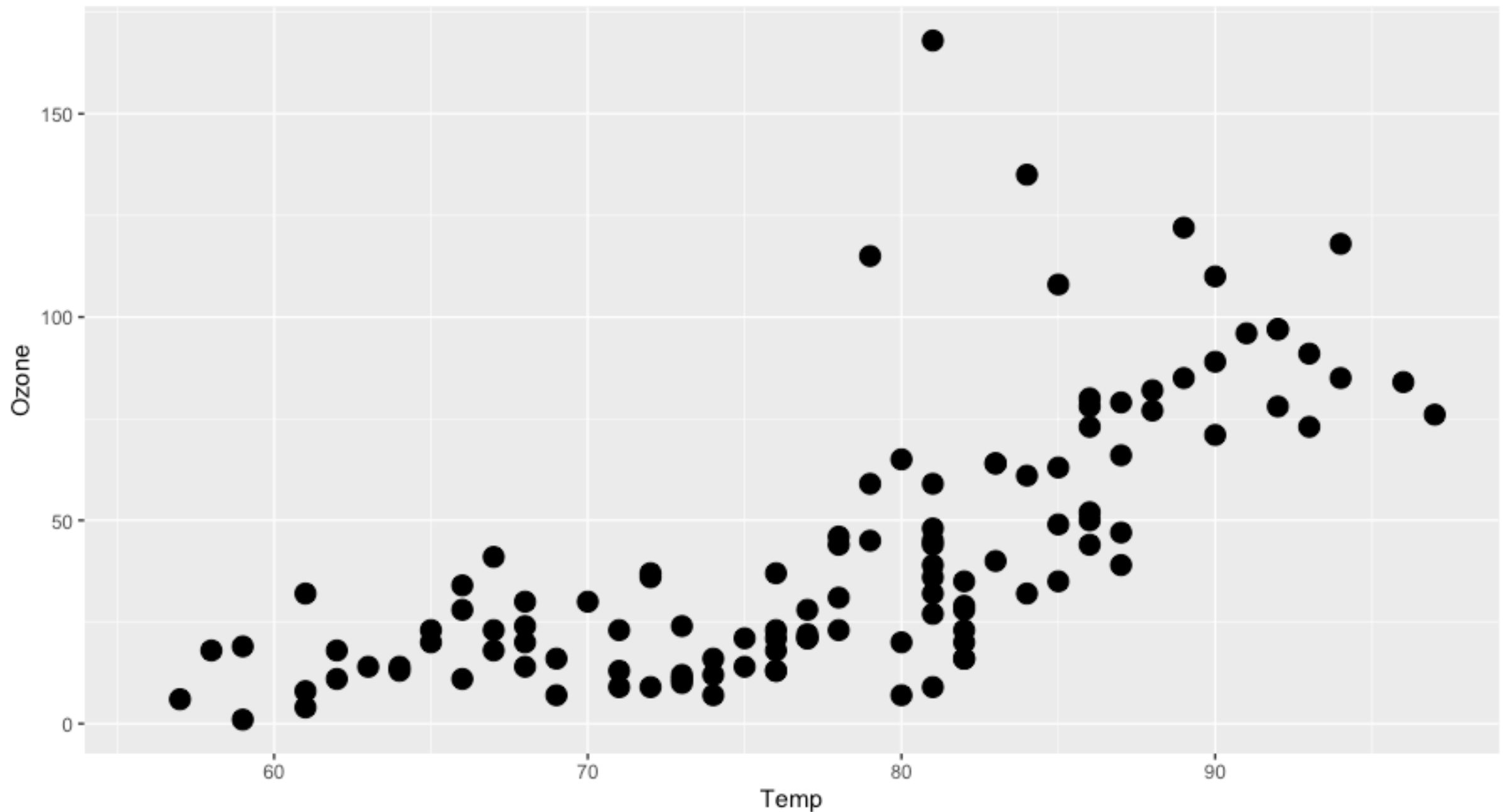
- One package on CRAN that contains interesting and useful color palettes
- Three types of palettes: sequential, diverging, qualitative
- Can be invoked in ggplot2 using the **scale_color_brewer()** function
- Designed to be used with discrete data

Colorbrewer Palettes



Using Color Palettes

```
airquality %>%  
  ggplot(aes(Temp, Ozone)) +  
  geom_point(size = 4)
```



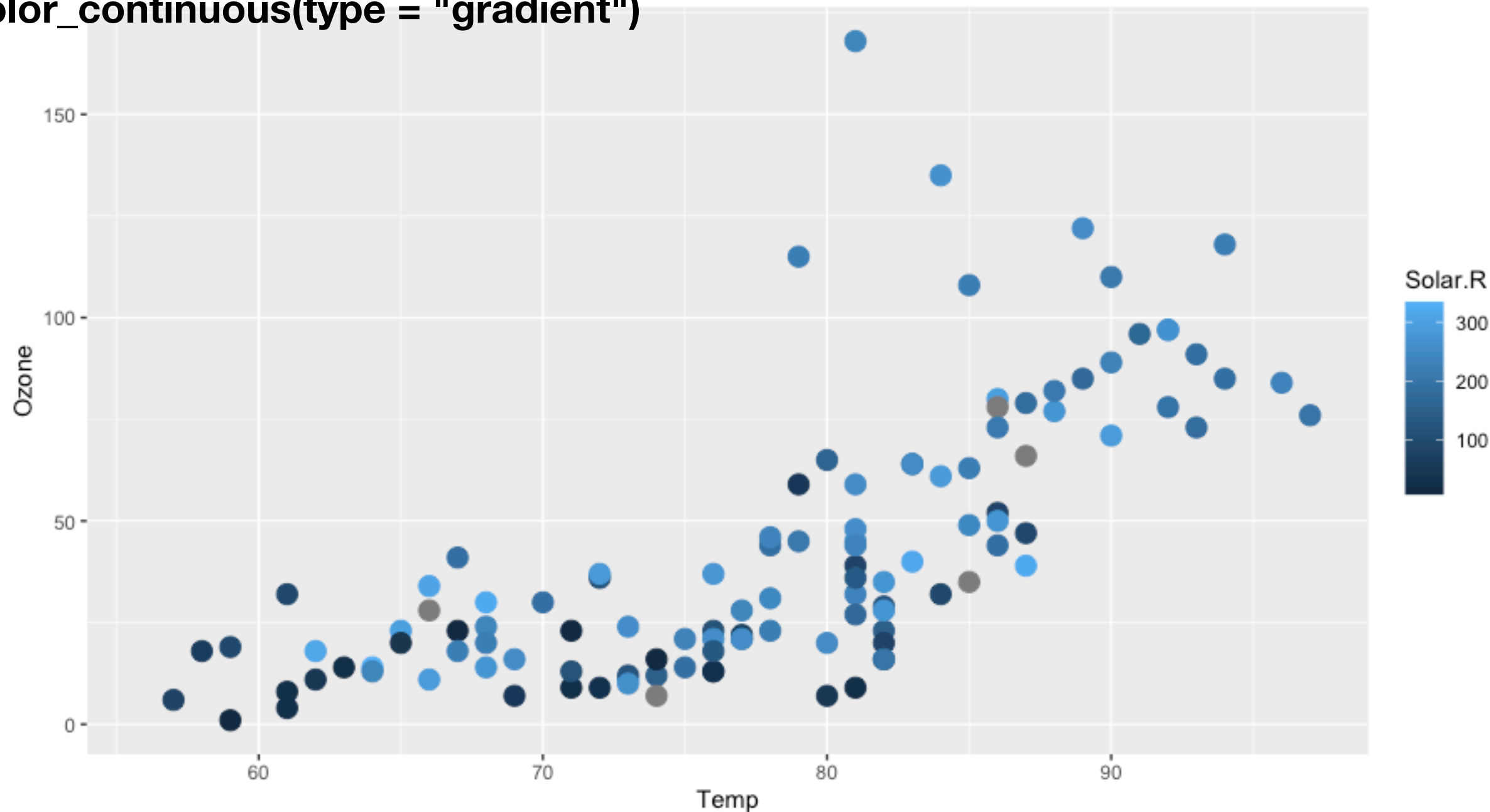
Continuous / Sequential Color Palettes

airquality %>%

ggplot(aes(Temp, Ozone)) +

geom_point(aes(color = Solar.R), size = 4) +

scale_color_continuous(type = "gradient")



Continuous / Sequential Color Palettes

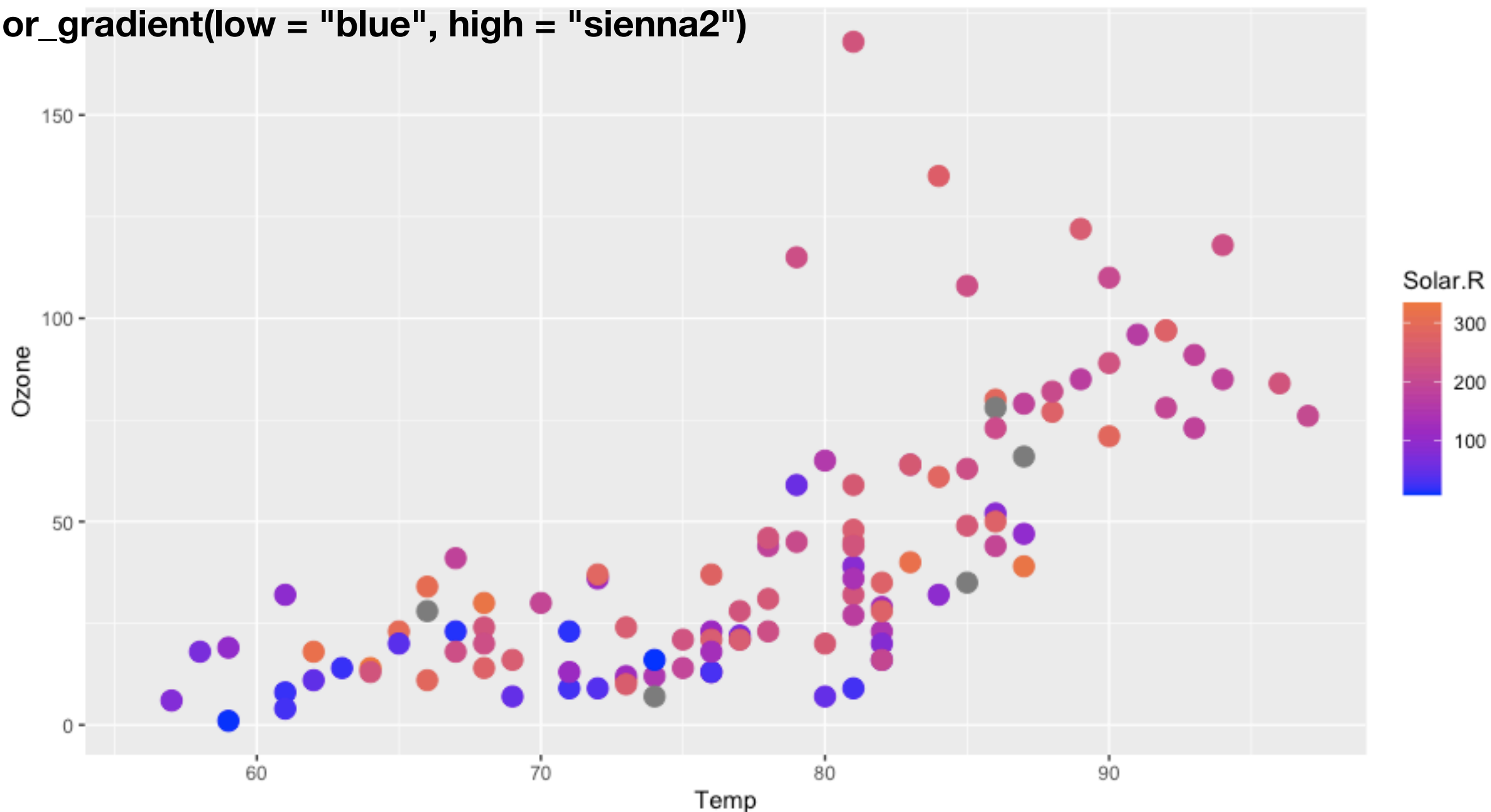
```
airquality %>%
```

```
  ggplot(aes(Temp, Ozone)) +
```

```
  geom_point(aes(color = Solar.R), size = 4) +
```

```
  scale_color_gradient(low = "blue", high = "sienna2")
```

The `colors()` function lists all
colors that R recognizes by name



Traversing Color Space



Qualitative / Categorical Color Palettes

```
airquality %>%
```

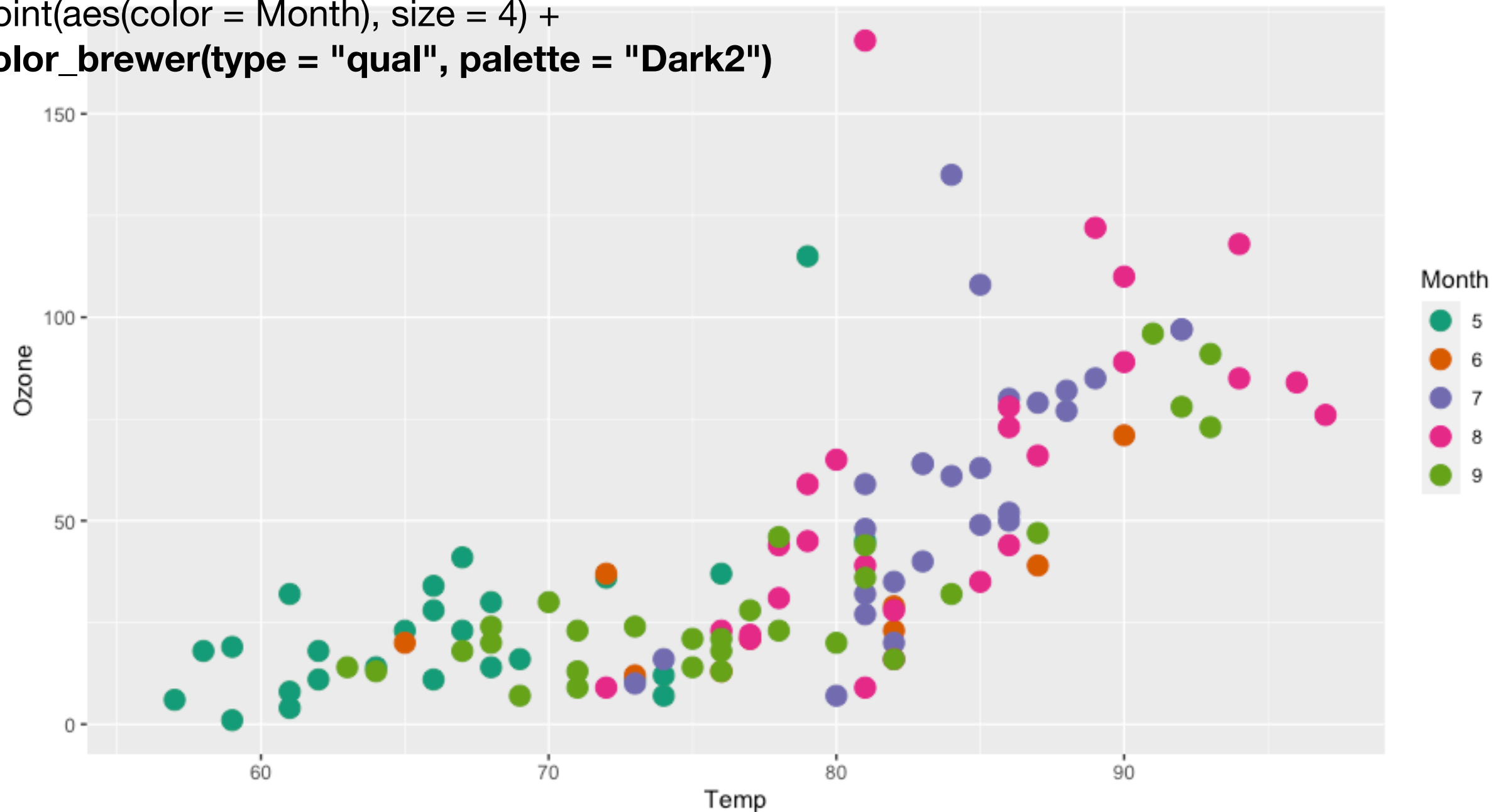
```
  mutate(Month = factor(Month)) %>%
```

```
  ggplot(aes(Temp, Ozone)) +
```

```
  geom_point(aes(color = Month), size = 4) +
```

```
  scale_color_brewer(type = "qual", palette = "Dark2")
```

Convert 'Month' into a categorical variable



Viridis Palette

- Designed to maximize color range
- Perceptually uniform
- Robust to various forms of color-blindness
- Can be mapped to monochrome (b/w) environments
- Reasonably attractive
- Useful in a wide range of scenarios

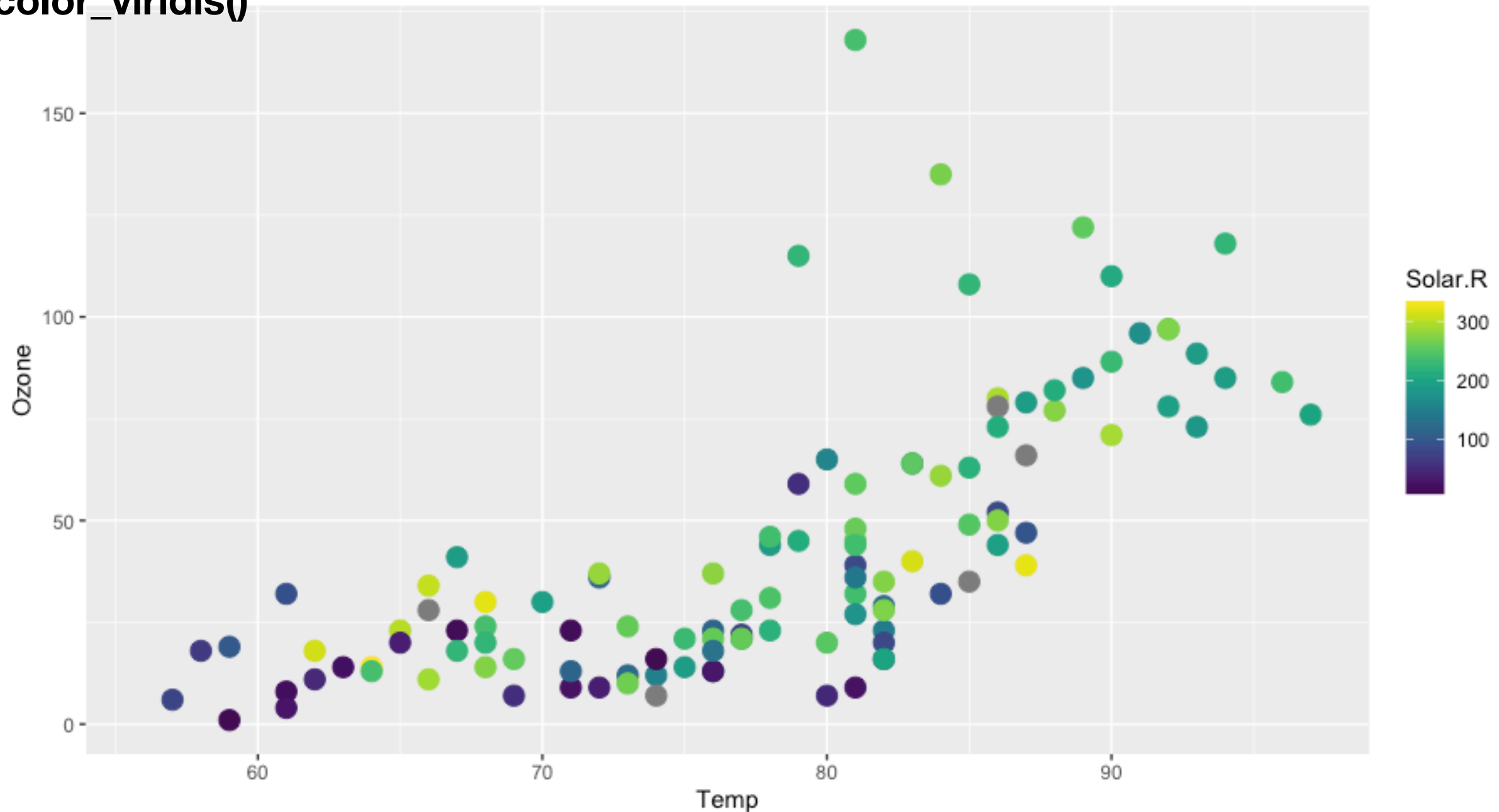
Viridis Palette

airquality %>%

ggplot(aes(Temp, Ozone)) +

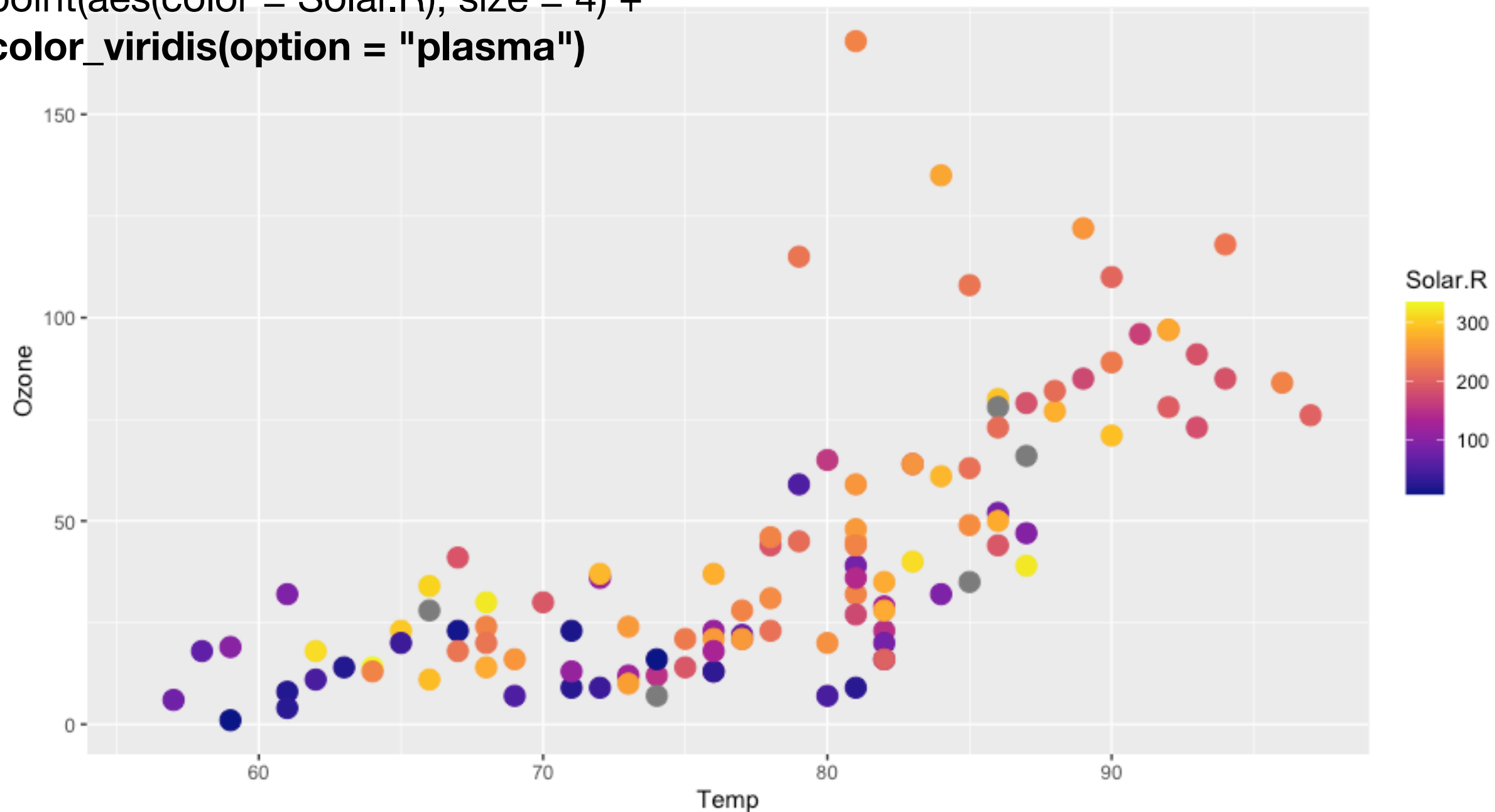
geom_point(aes(color = Solar.R), size = 4) +

scale_color_viridis()



Viridis Palette

```
airquality %>%  
  ggplot(aes(Temp, Ozone)) +  
  geom_point(aes(color = Solar.R), size = 4) +  
  scale_color_viridis(option = "plasma")
```



Customizing Color

airquality %>%

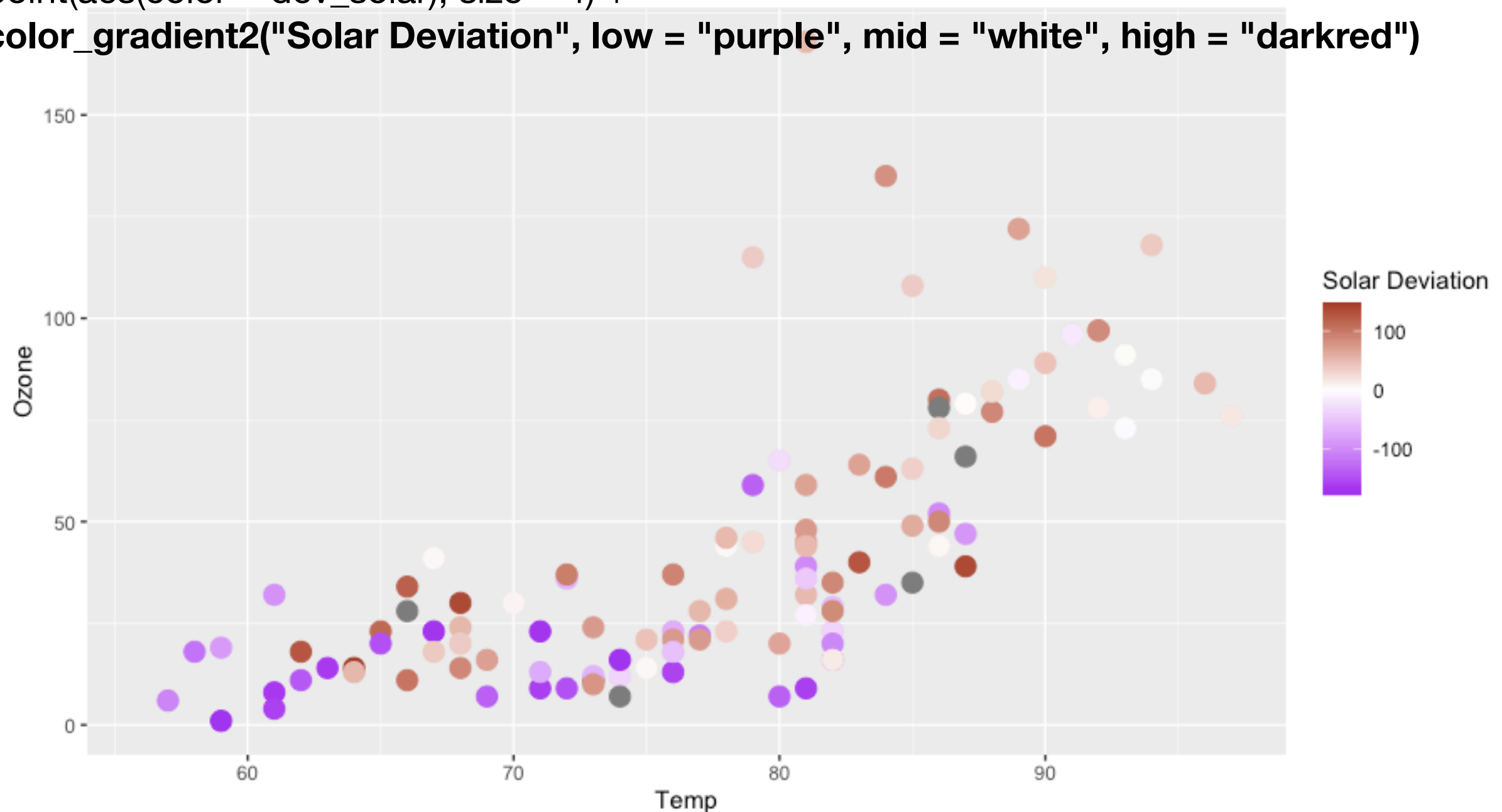
```
mutate(dev_solar = Solar.R - mean(Solar.R, na.rm = TRUE)) %>%
```

```
ggplot(aes(Temp, Ozone), alpha = 1/5) +
```

```
geom_point(aes(color = dev_solar), size = 4) +
```

```
scale_color_gradient2("Solar Deviation", low = "purple", mid = "white", high = "darkred")
```

Create deviation
variable



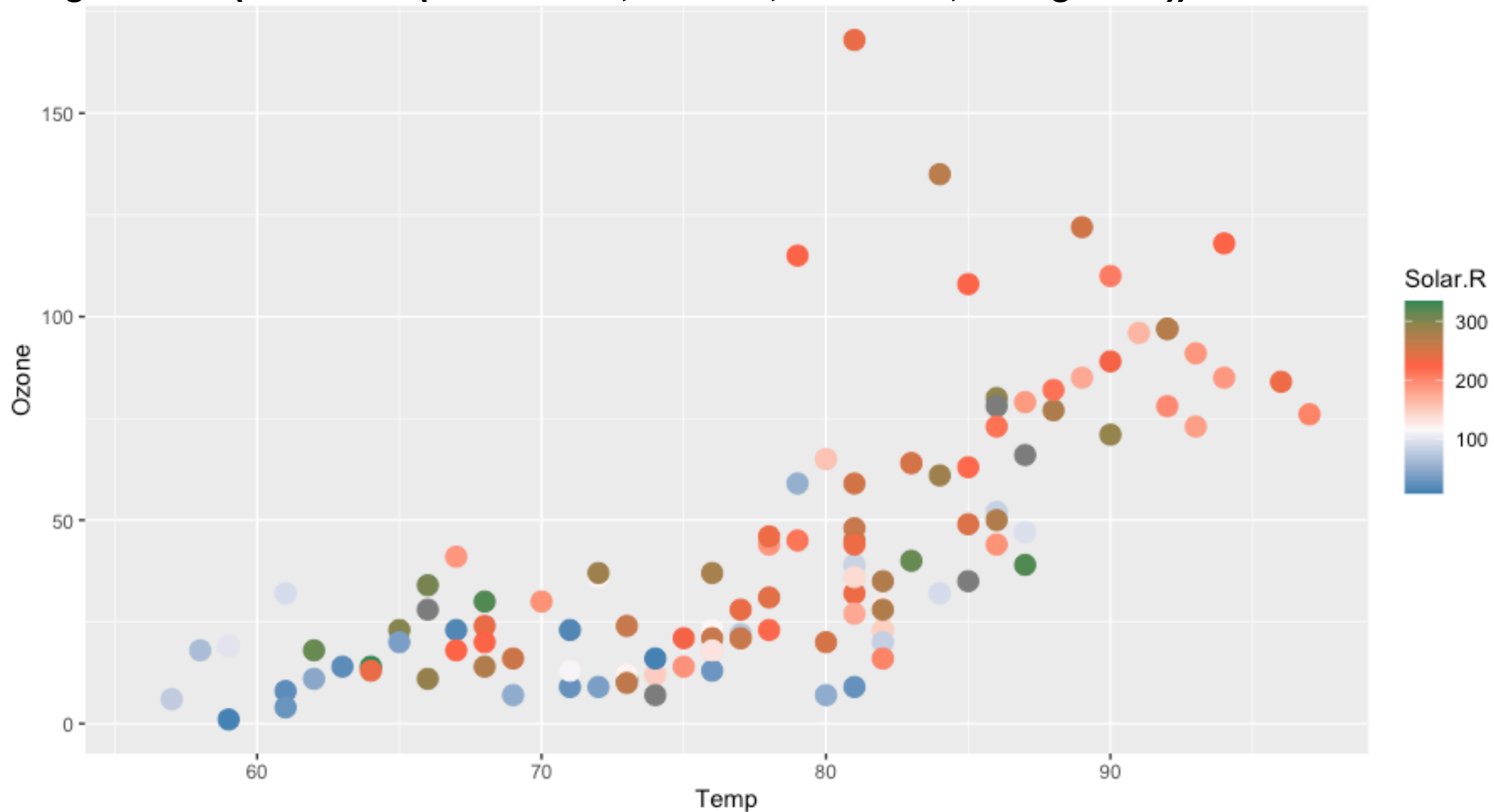
Customizing Color

airquality %>%

ggplot(aes(Temp, Ozone)) +

geom_point(aes(color = Solar.R), size = 4) +

scale_color_gradientn(colors = c("steelblue", "snow", "tomato", "seagreen"))



Customizing Color

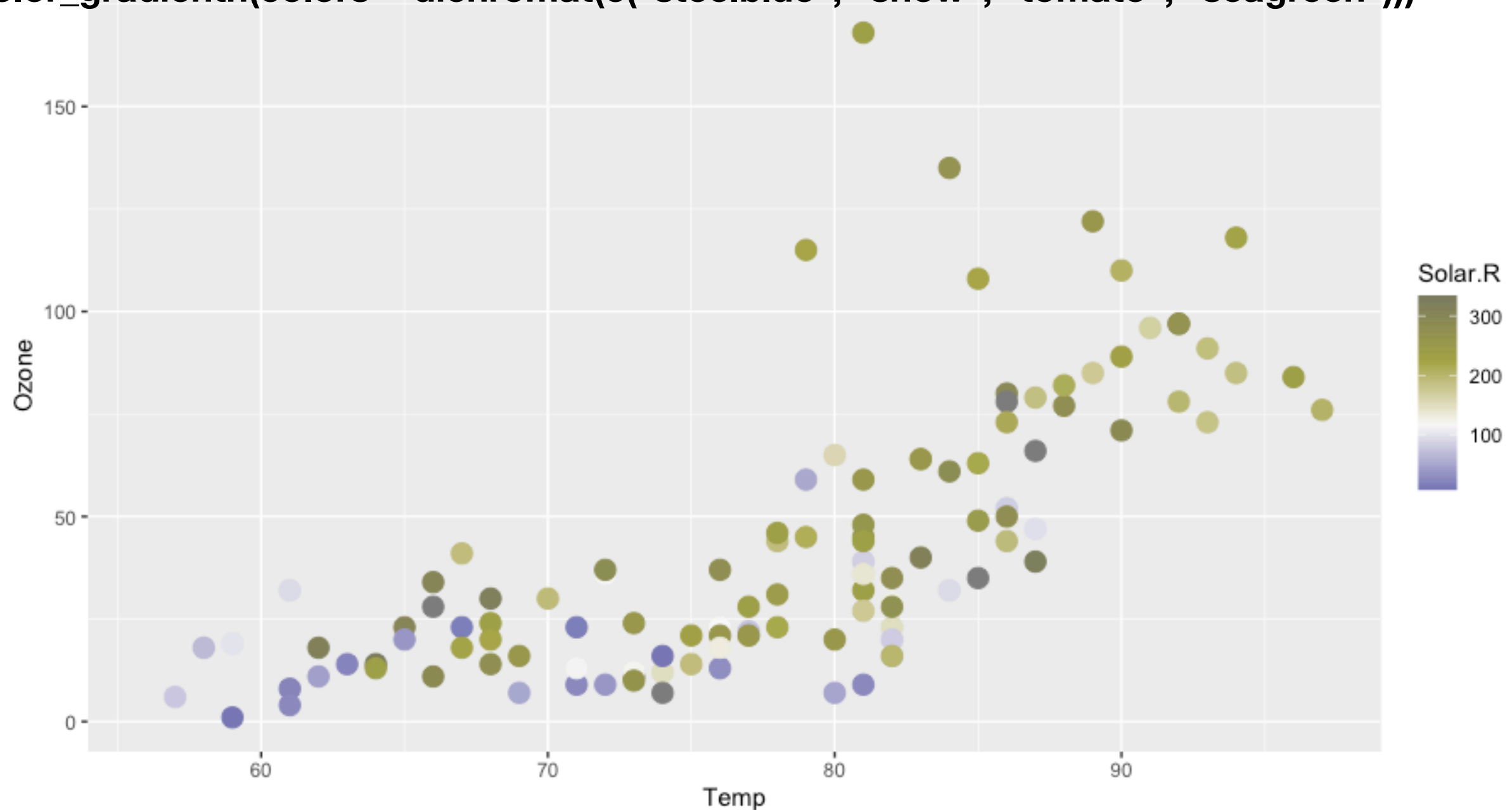
```
library(dichromat)
```

```
airquality %>%
```

```
  ggplot(aes(Temp, Ozone)) +
```

```
  geom_point(aes(color = Solar.R), size = 4) +
```

```
  scale_color_gradientn(colors = dichromat(c("steelblue", "snow", "tomato", "seagreen"))))
```



Color Summary

- Colors can be very useful for conveying quantitative or qualitative information in data graphics
- Care must be taken to ensure that color contrasts can be perceived by the audience (can check with dichromat package)
- Built in ggplot2 color palettes and Viridis palettes are often reasonable choices and well-calibrated
- Some customization can be done with `scale_color_gradient()` functions