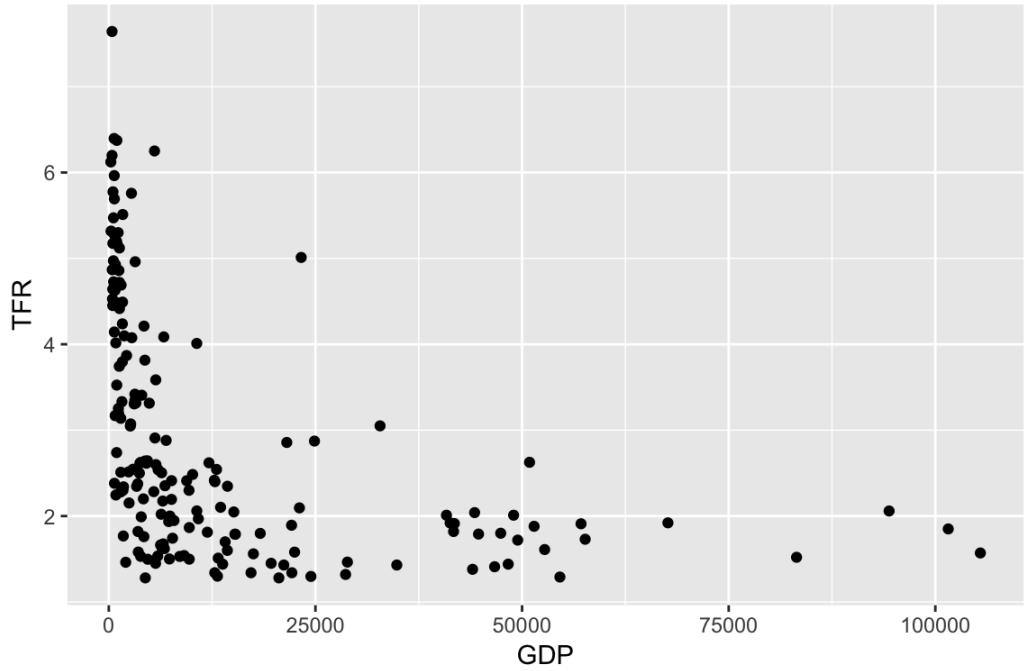
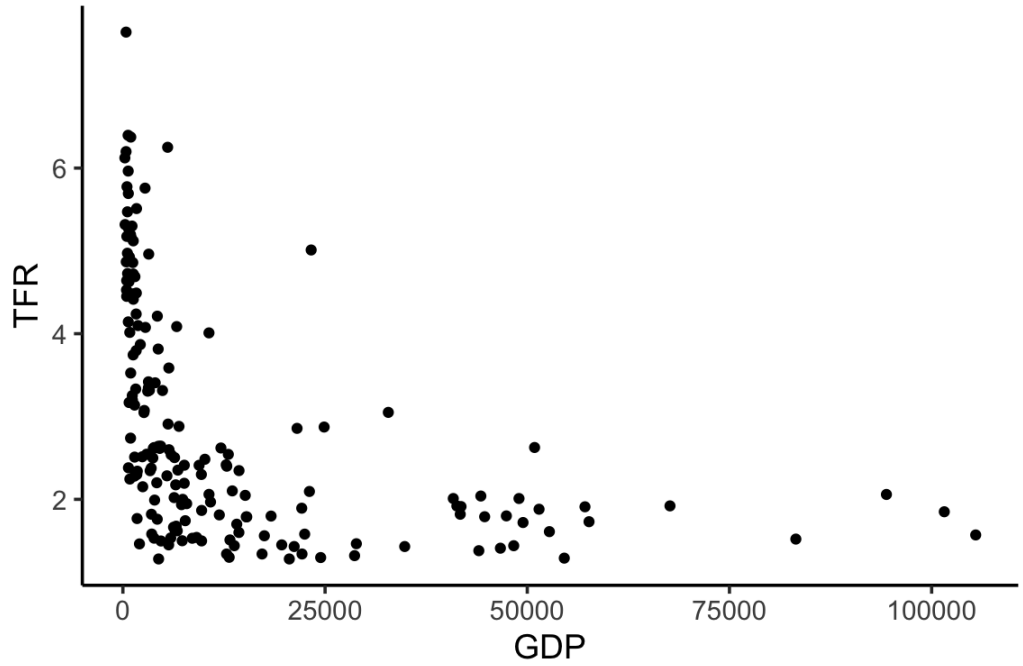
Total Fertility Rate vs. Gross Domestic Product



(easier to read individual values with gridlines)

Total Fertility Rate vs. Gross Domestic Product



easier to detect spatial patterns without gridlines

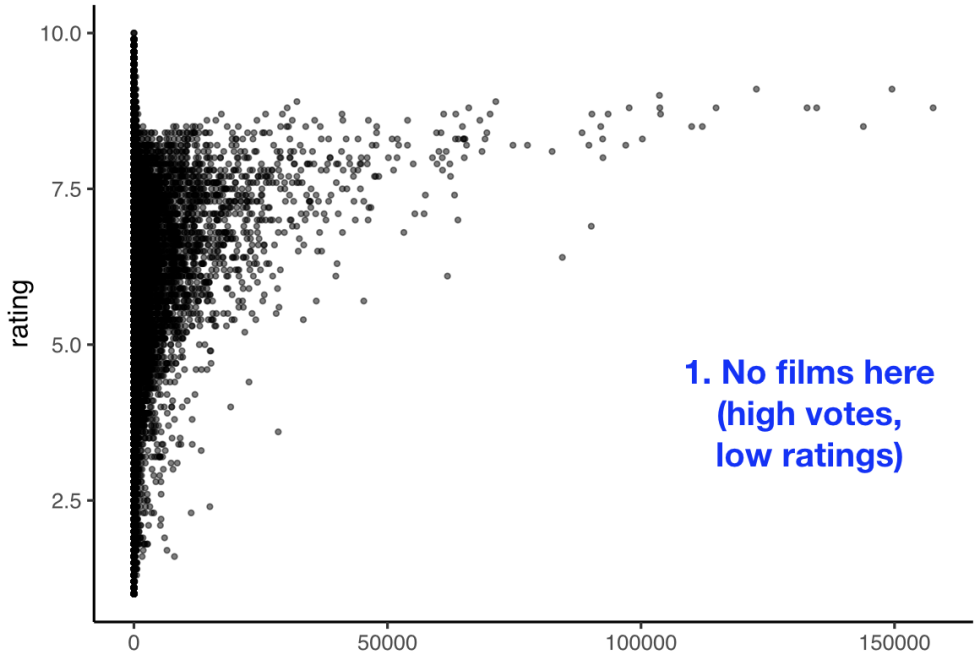
What features might be visible in scatterplots? (p. 77)

- Causal relationships

What features might be visible in scatterplots? (p. 77)

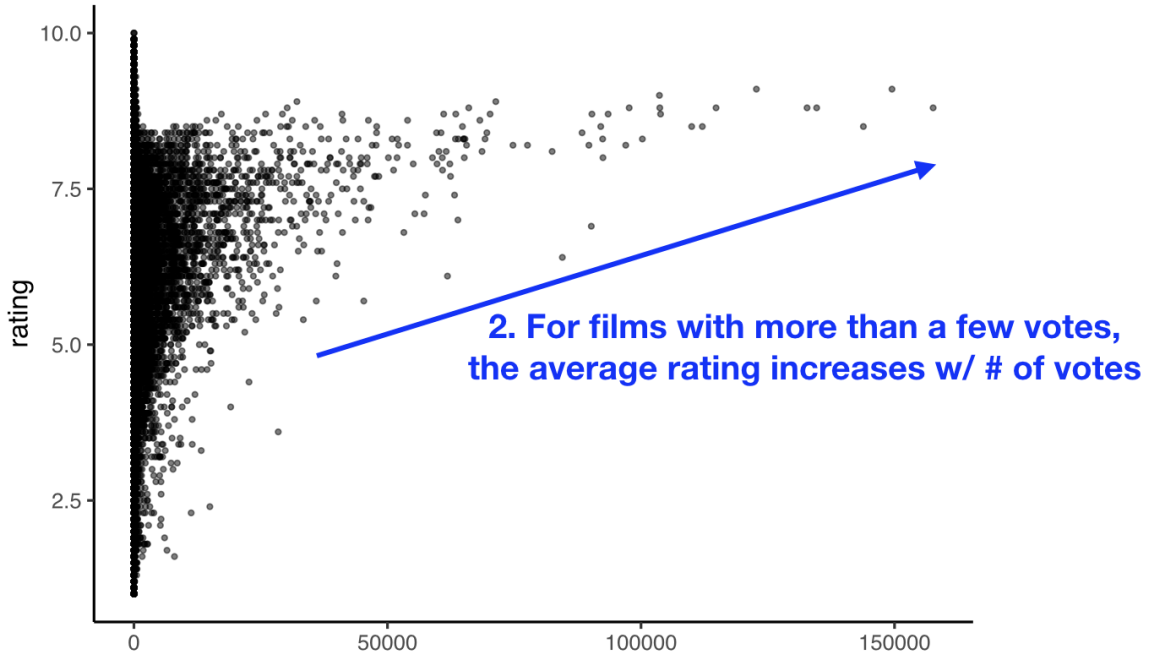
- ~~Causal relationships~~ (correlation ≠ causation, but still use y-axis for what appears to be the dependent variable)
- Associations
  describe what you see
- Outliers
- Clusters
- Gaps
- Barriers (boundaries)
- Conditional relationships
  (different relationships for different intervals of x)

Movie Ratings (GDwR, p. 82)

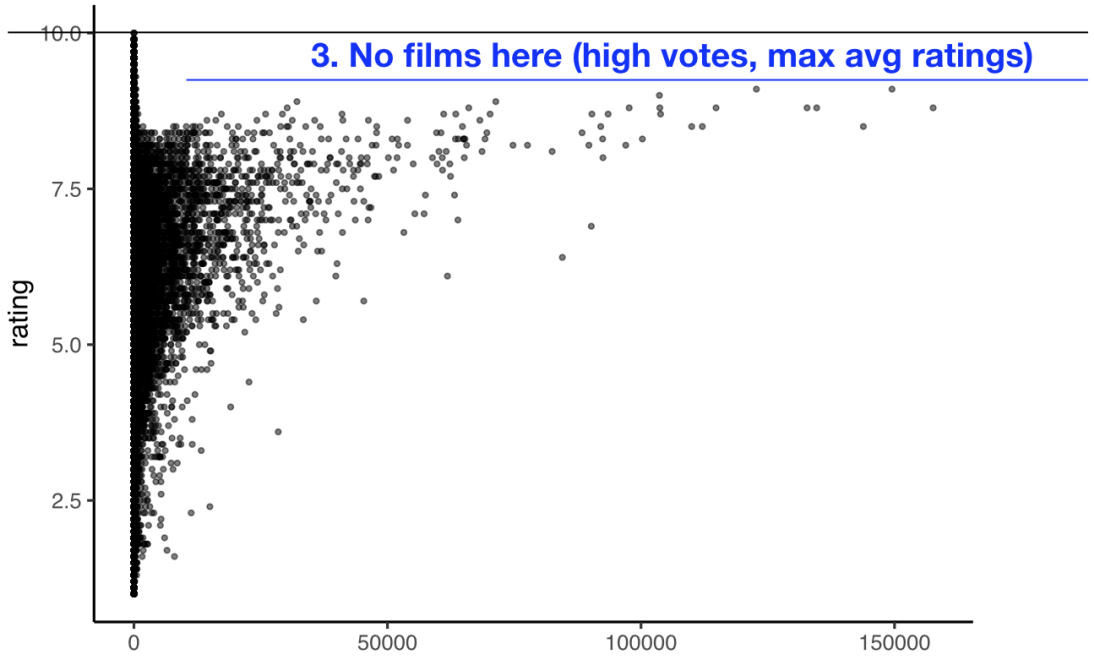


Source: Antony Unwin, *Graphical Data Analysis with R*

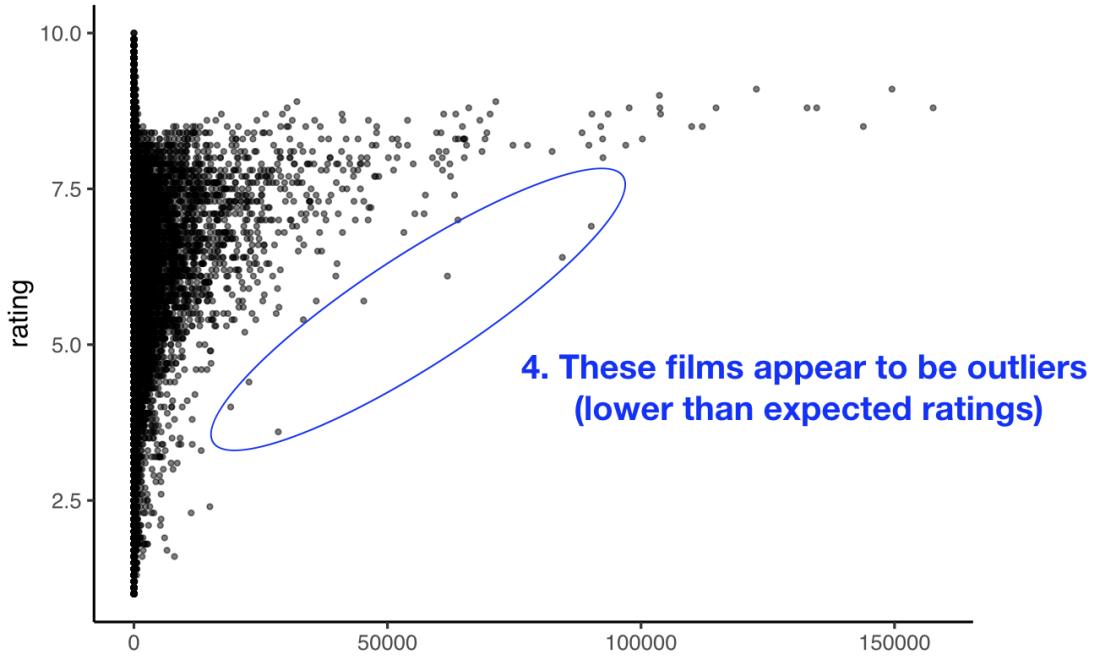
Movie Ratings (GDwR, p. 82)



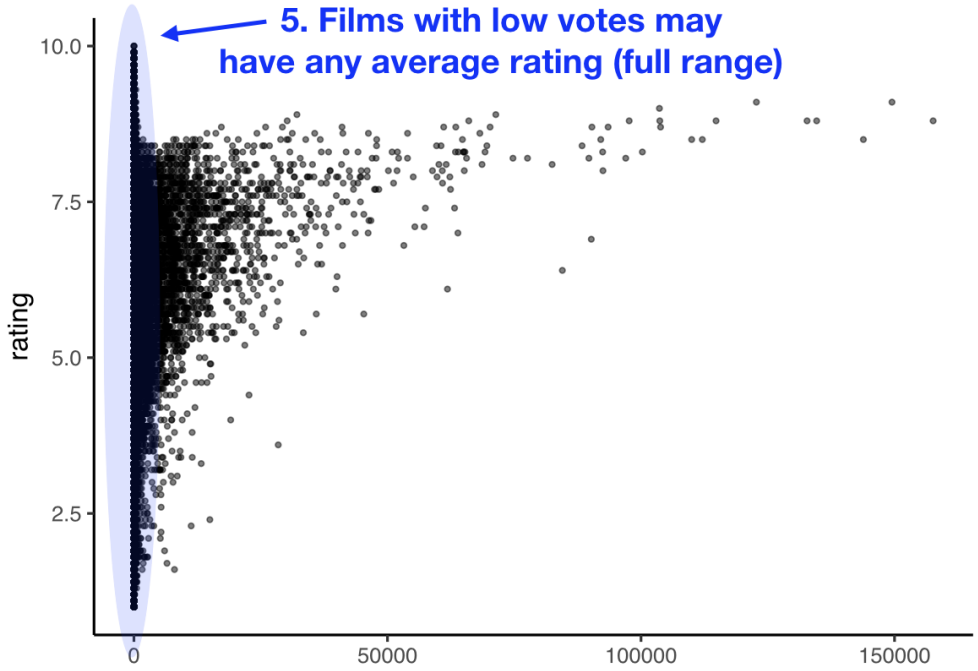
Movie Ratings (GDwR, p. 82)



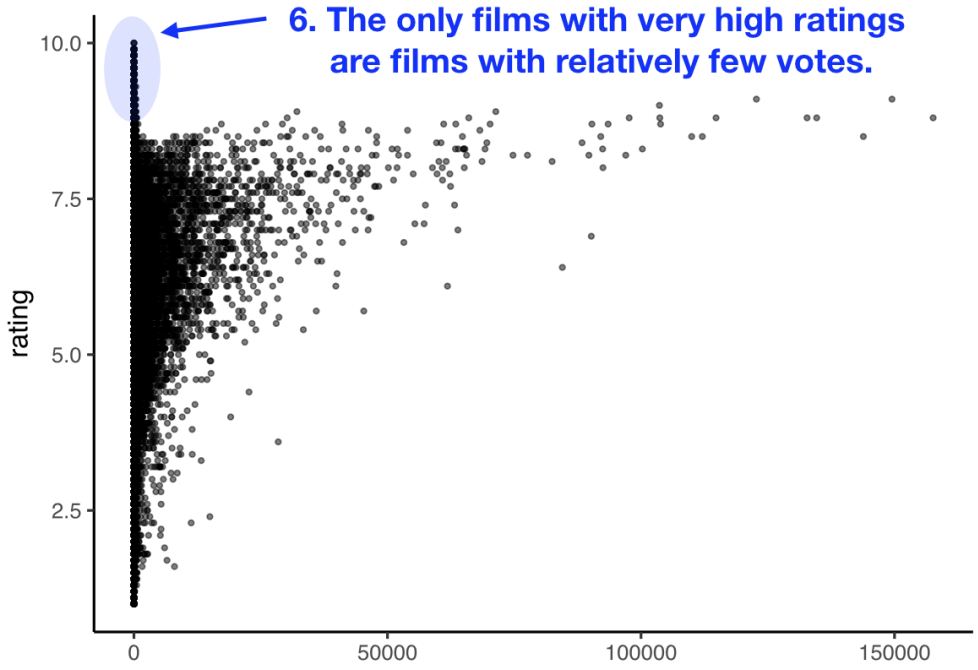
Movie Ratings (GDwR, p. 82)



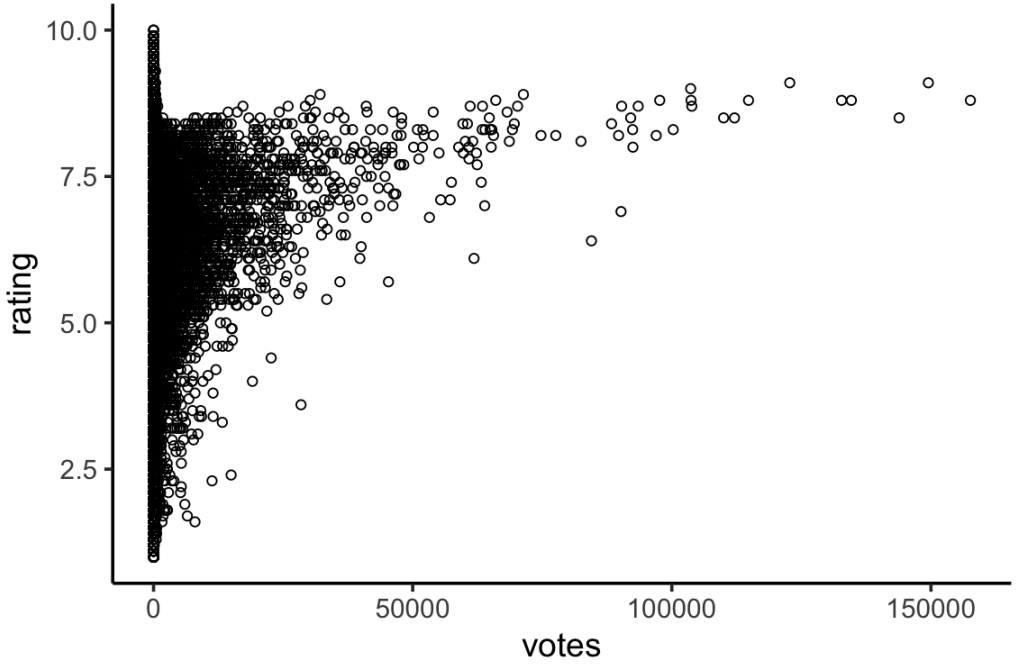
Movie Ratings (GDwR, p. 82)



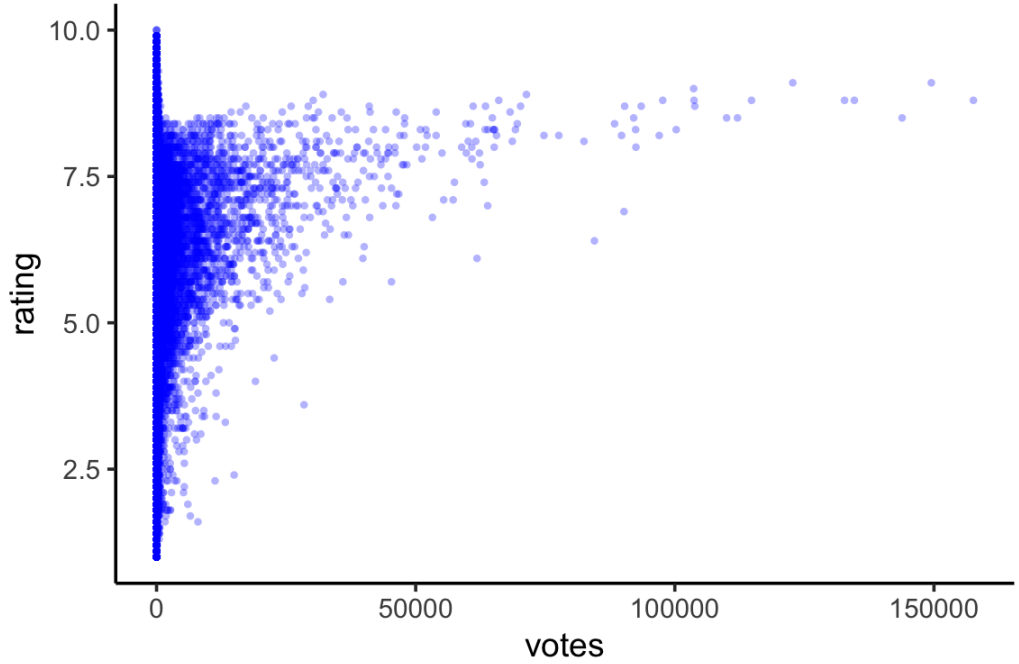
Movie Ratings (GDwR, p. 82)



Open circles

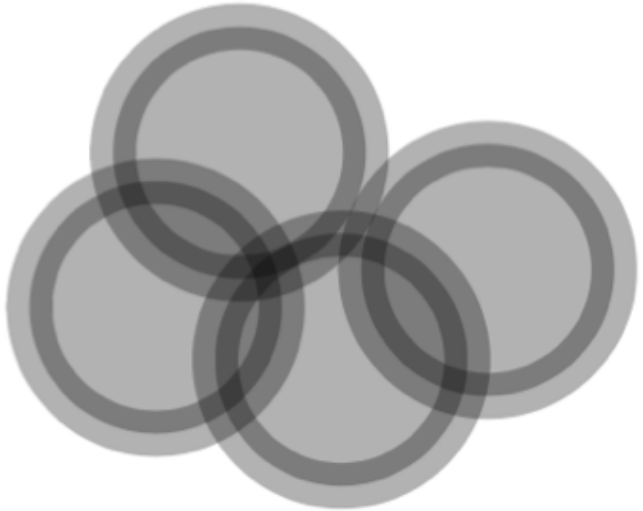


Alpha blending

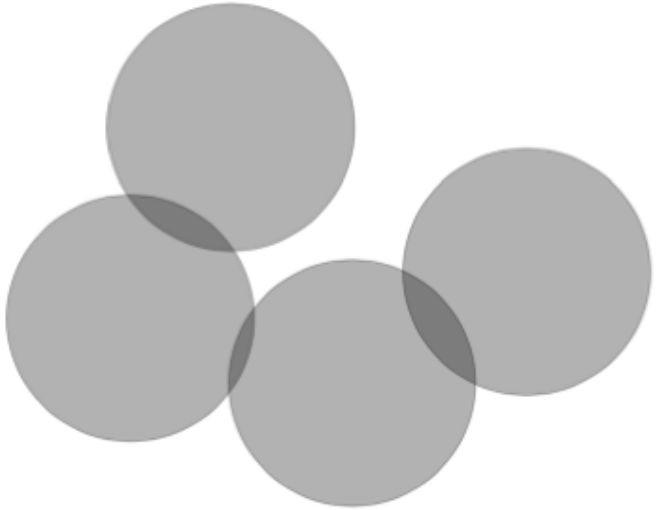


Alpha blending tip (ggplot2)

```
geom_point(alpha = .3)
```



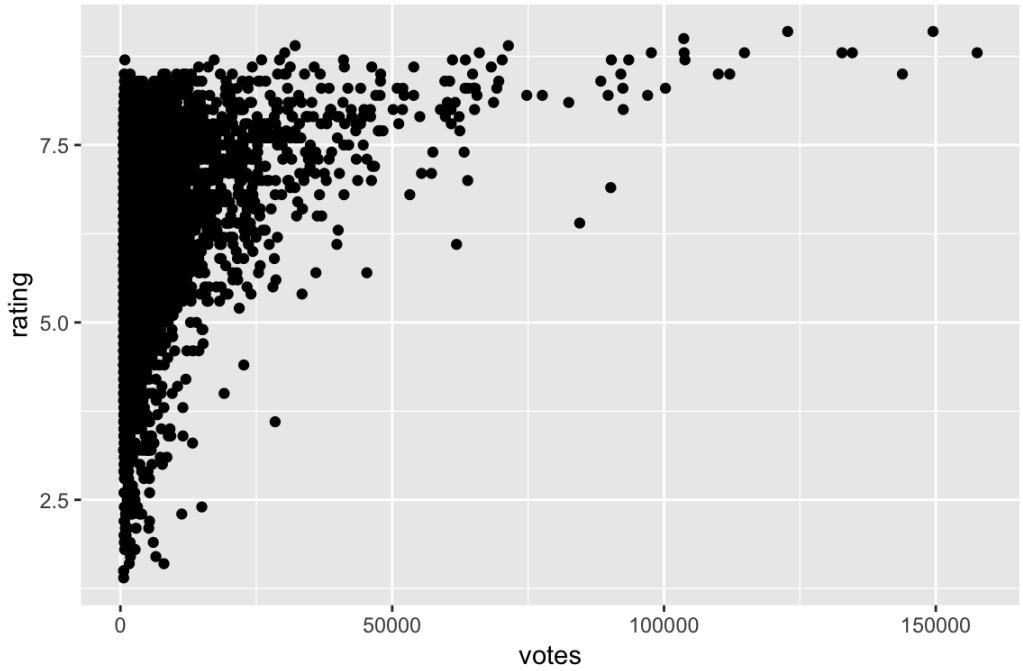
```
geom_point(alpha = .3,
           stroke = 0)
```



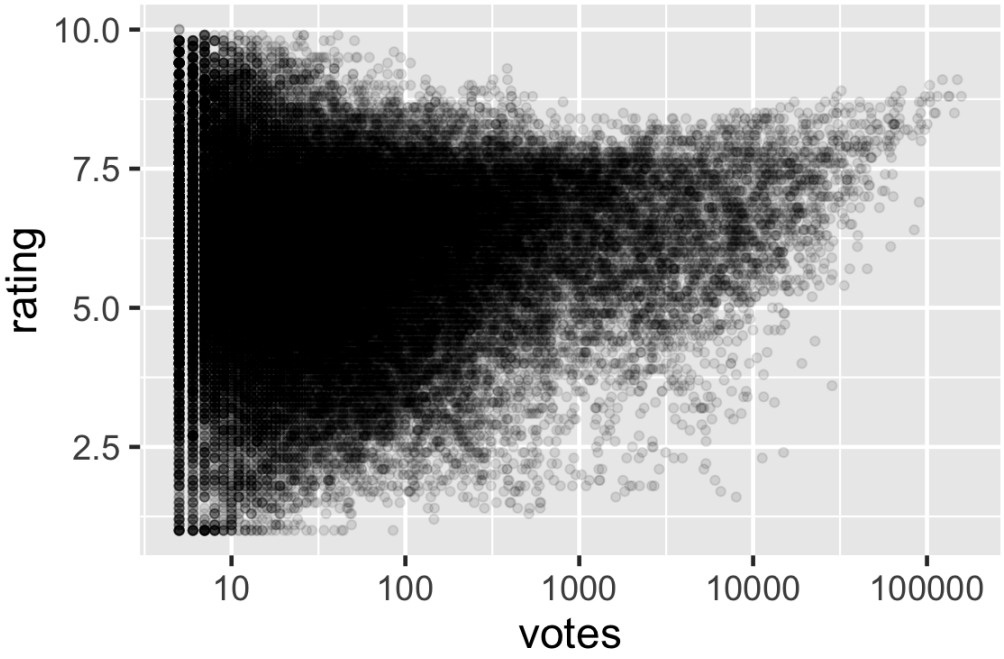
Strategies

- Use techniques to deal with overplotting (alpha blending, open circles, etc.)
- Don’t plot all points (remove outliers, subset data, sample data)
- Transform to log scale
- Heatmaps (bin counts or density estimates)
- Density contour lines
- Combination of above
- Multiple variables: scatterplot matrices

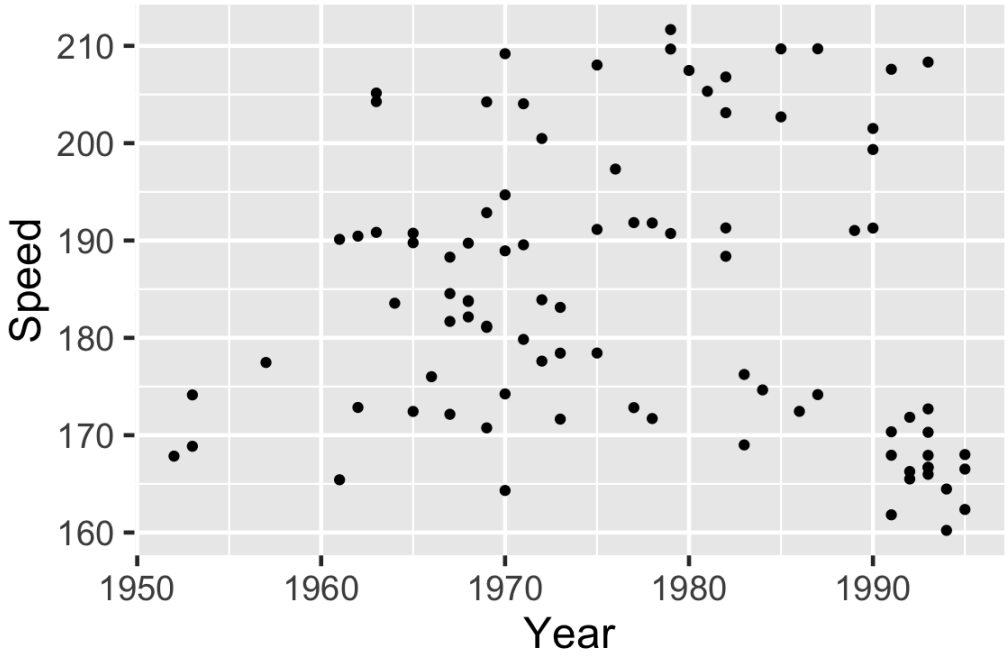
10% with highest number of votes



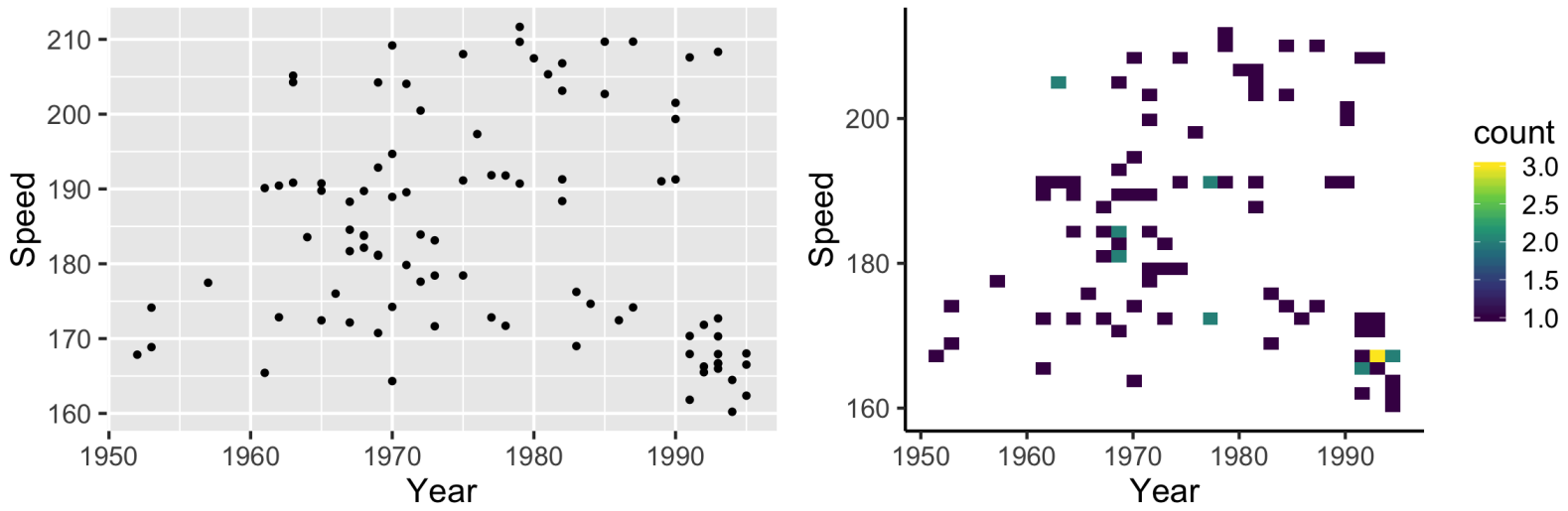
Log scale



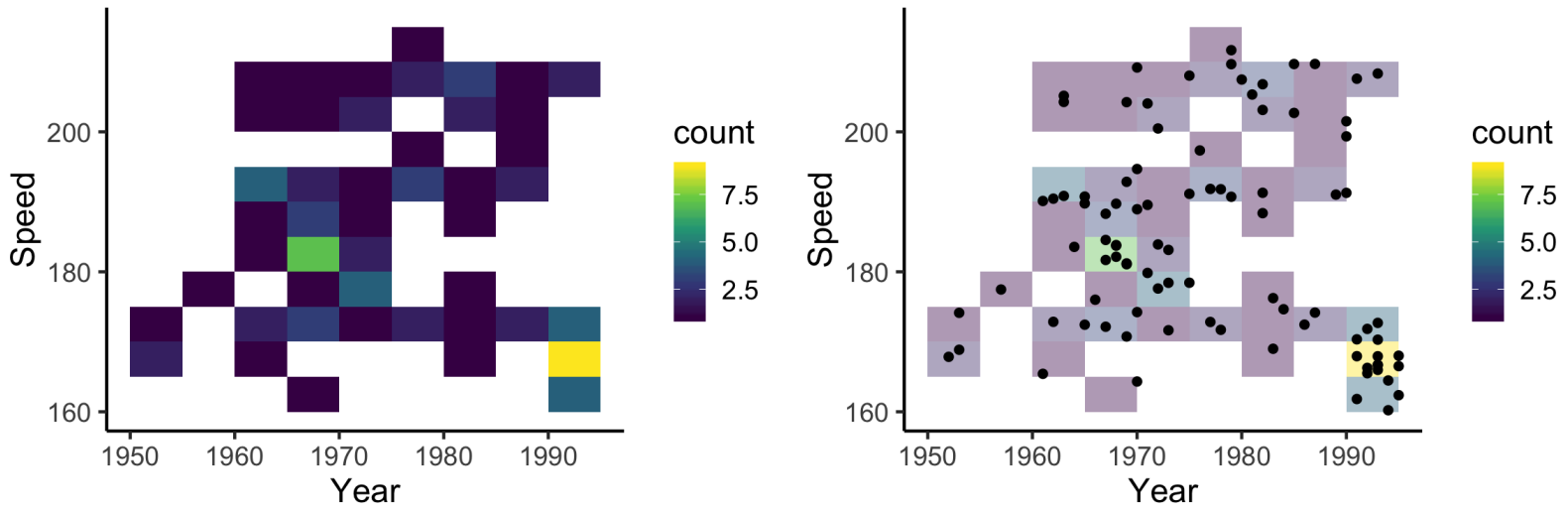
SpeedSki data (2011)



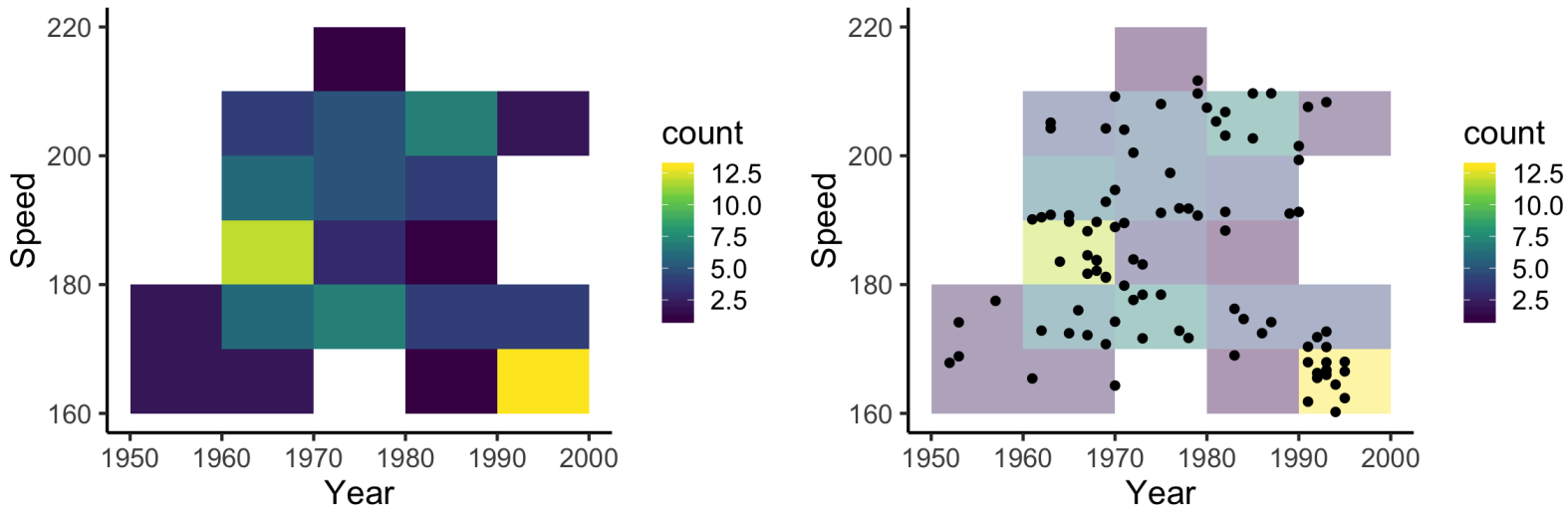
Square heatmap of bin counts (default: 30 bins)



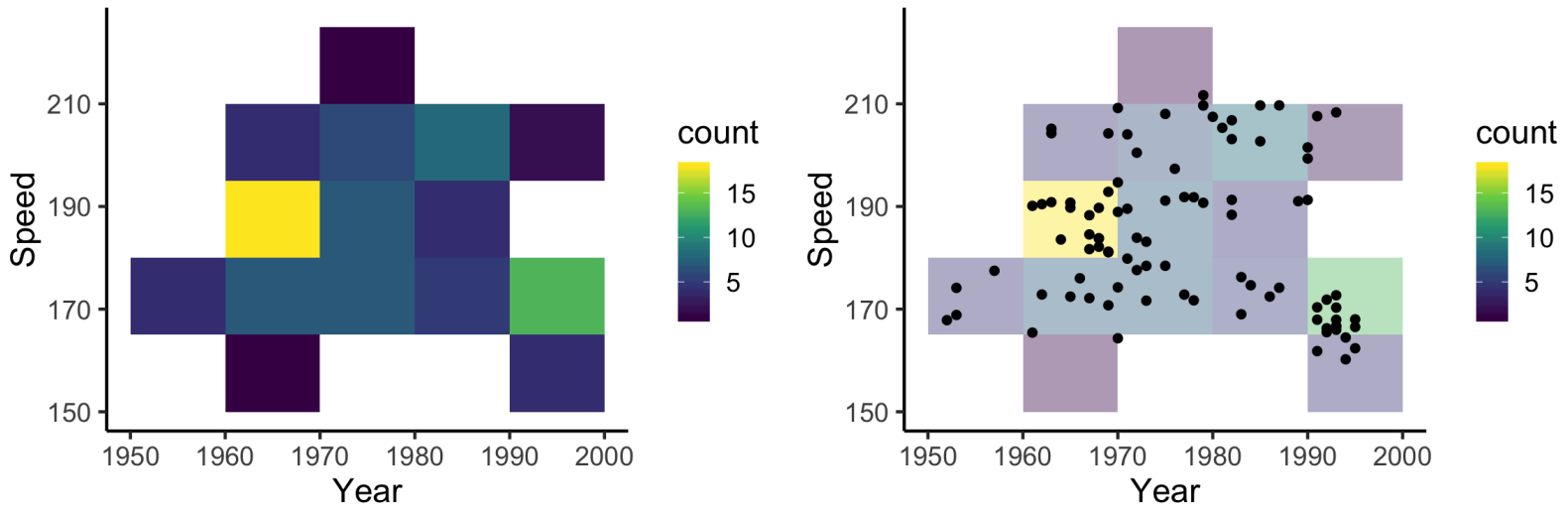
Square heatmap of bin counts (binwidth = 5)



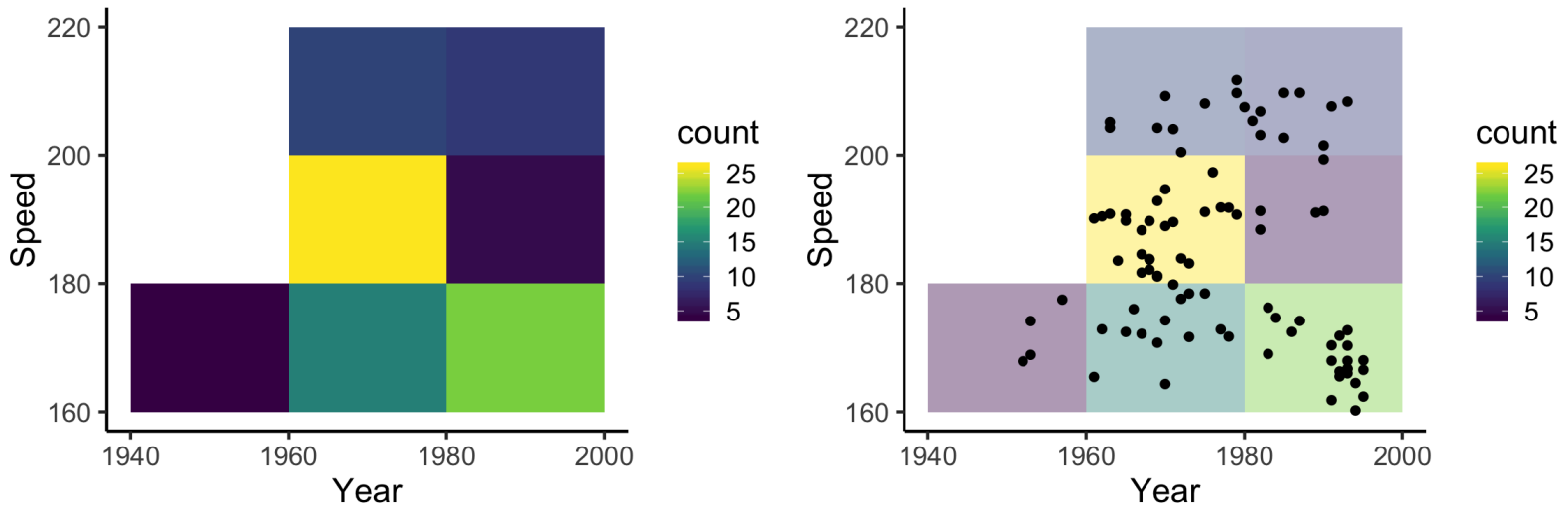
Square heatmap of bin counts (binwidth = 10)



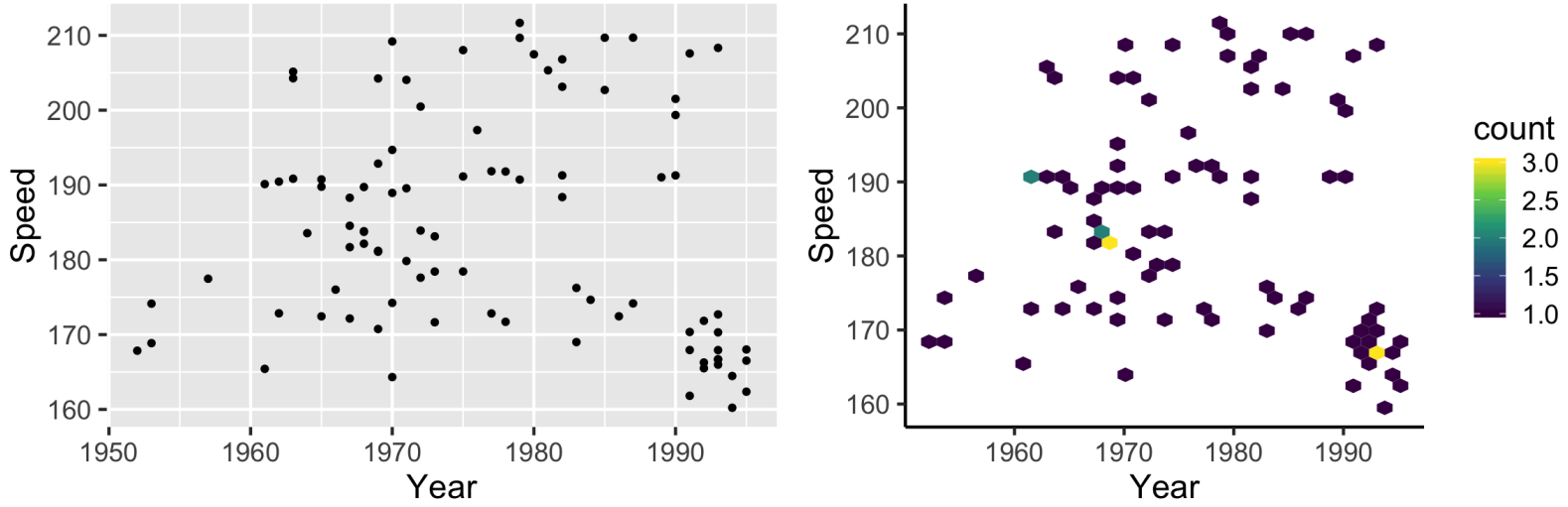
Square heatmap of bin counts (binwidth(x, y) = 10, 15)



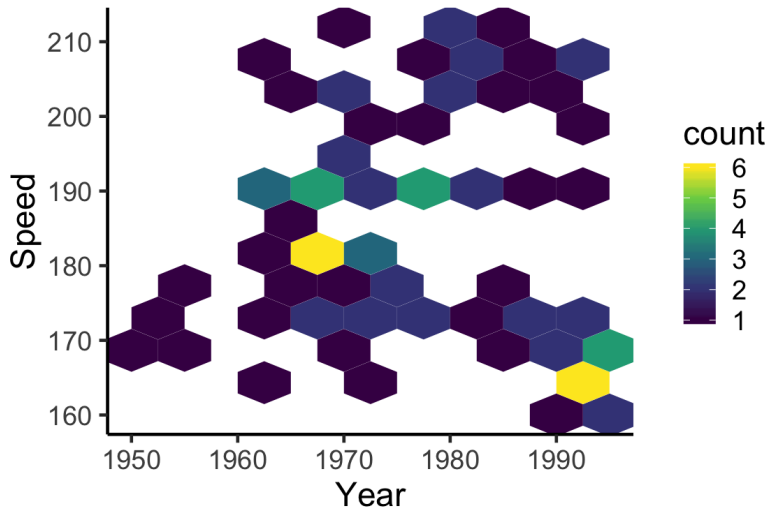
Square heatmap of bin counts (binwidth = 20)

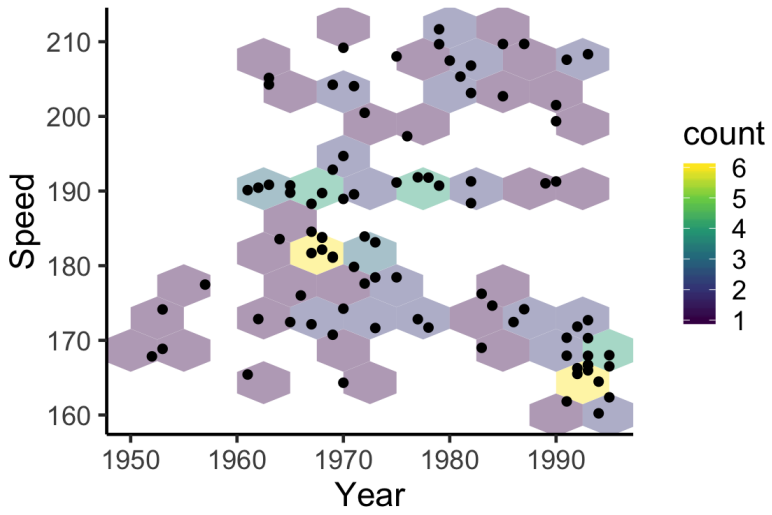


Hex heatmap of bin counts (default: 30 bins)

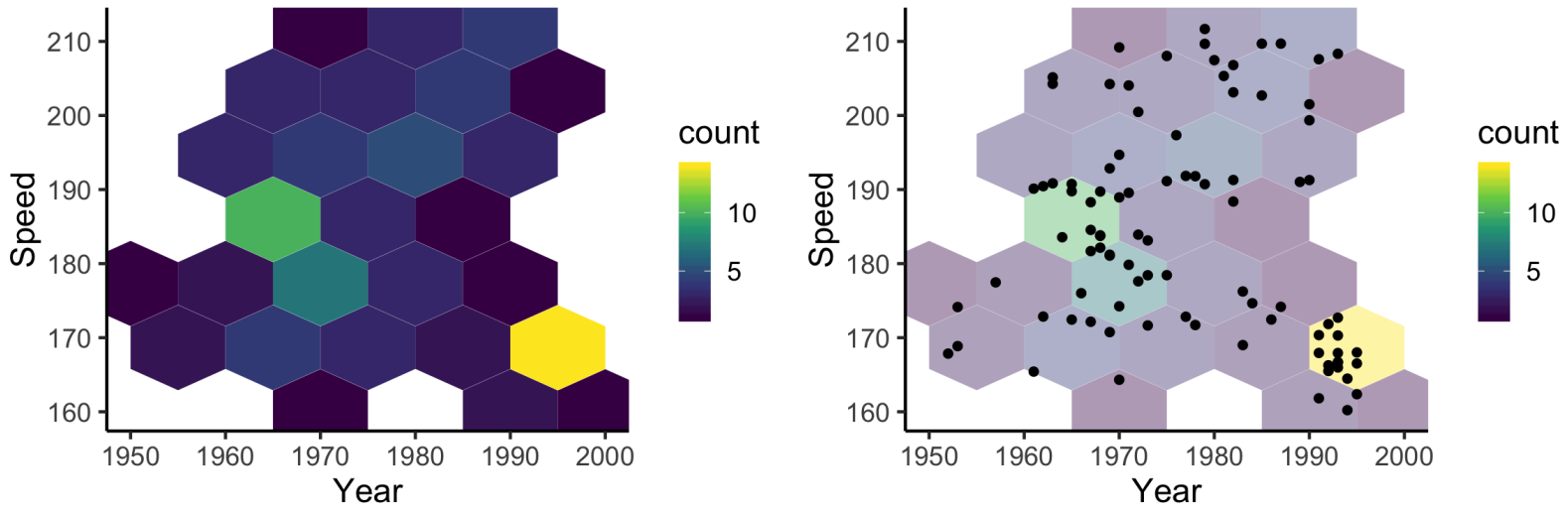


Hex heatmap of bin counts (binwidth = 5)





Hex heatmap of bin counts (binwidth = 10)



Hex heatmap of bin counts (binwidth(x, y) = 10, 15)