538 Calibration Chart with ggplot2

Larry Hignight

3/3/2020

MLB games, 2016-18

https://projects.fivethirtyeight.com/checking-our-work/
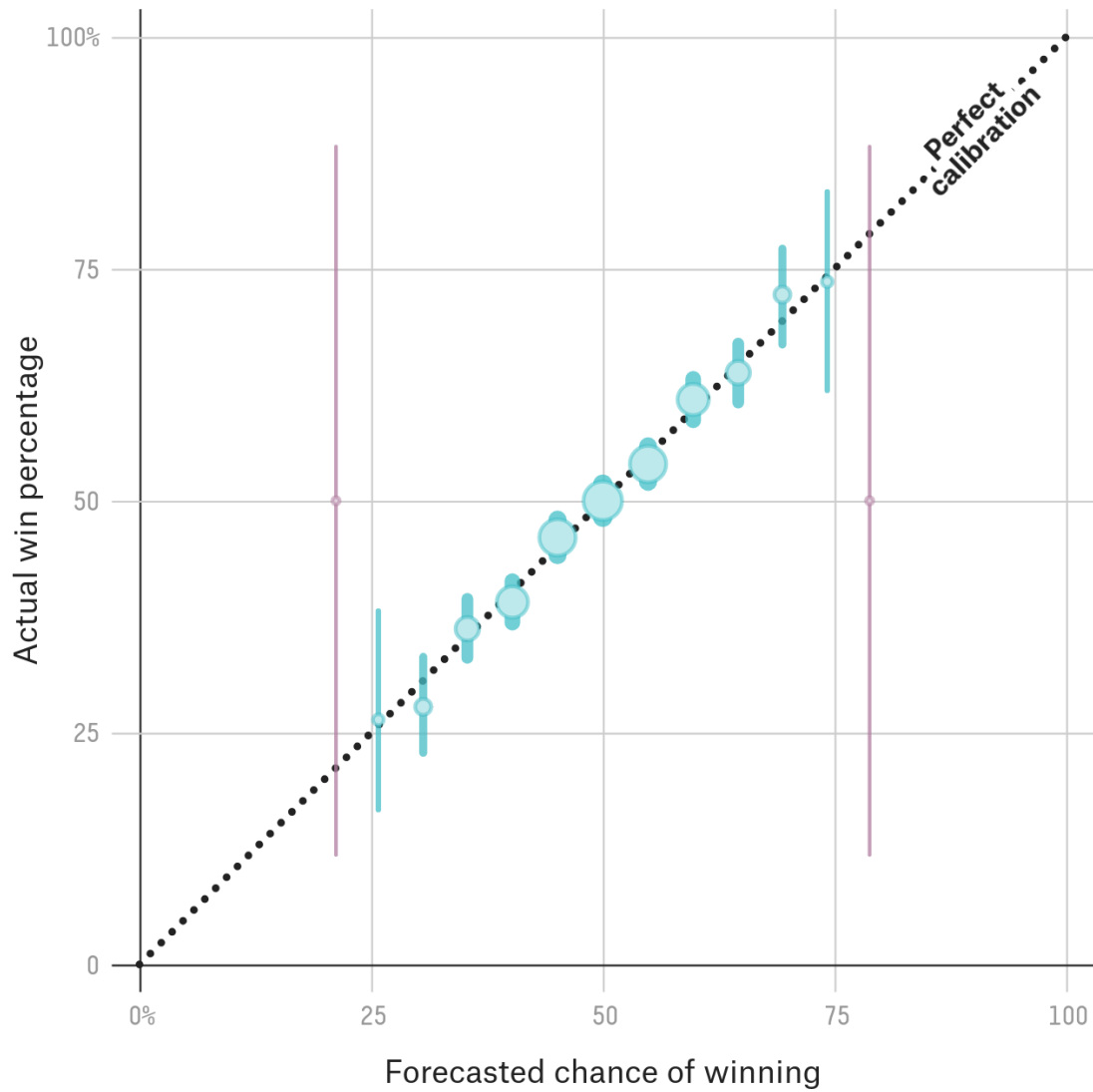
Chart Elements

Scatterplot Data:

- x - Forecasted win percentage
- y - Actual win percentage
- size - The number of games in a bin
- line - 95% confidence interval lines
- color
- point shape

Fixed Diagnal Line:

- Dotted black line
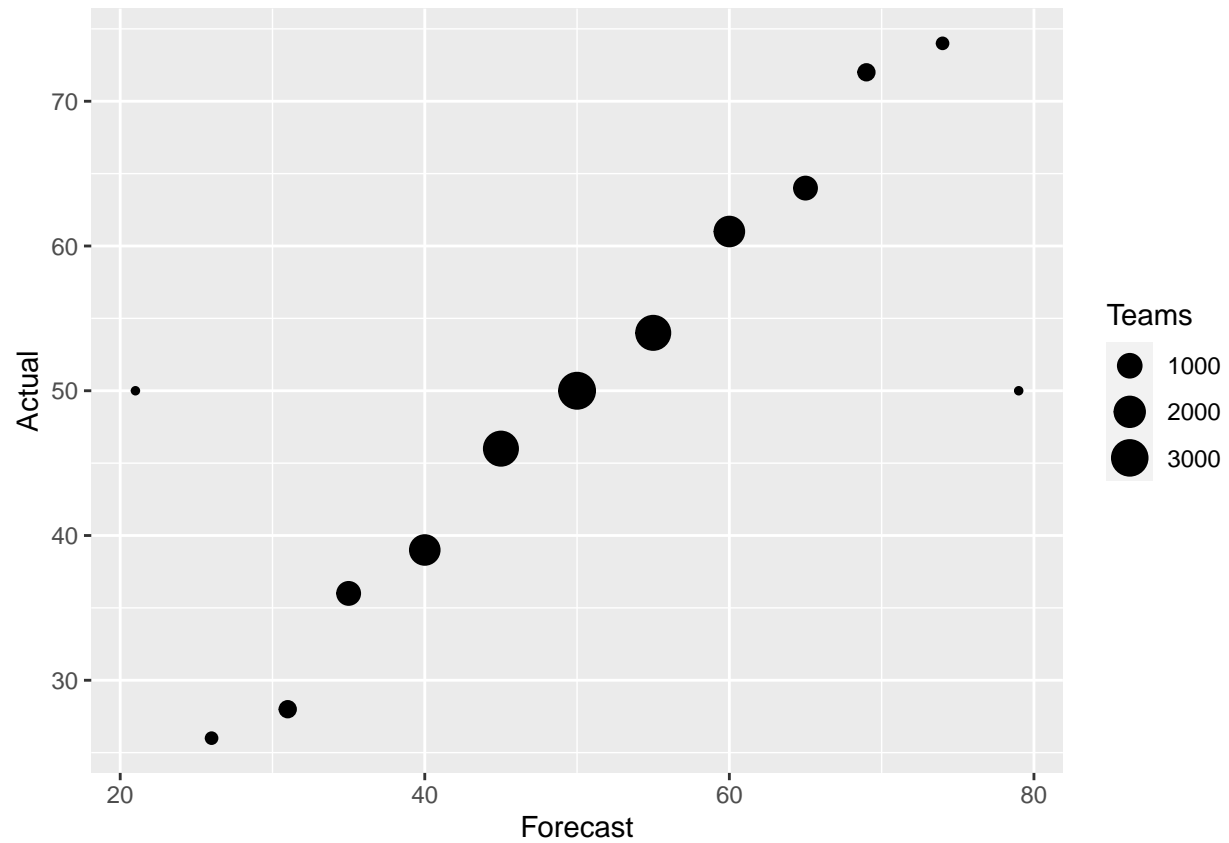- "Perfect Calibration" label

Labels:

- title, x label and y label

Data Import

```
x <- read.csv('mlb-data.csv')
head(x)
```

```
##   Teams Forecast Actual Interval
## 1     6       21     50     37.0
## 2    72       26     26     10.0
## 3   306       31     28      5.0
## 4   909       35     36      2.5
## 5  1868       40     39      2.5
## 6  2691       45     46      2.5
```
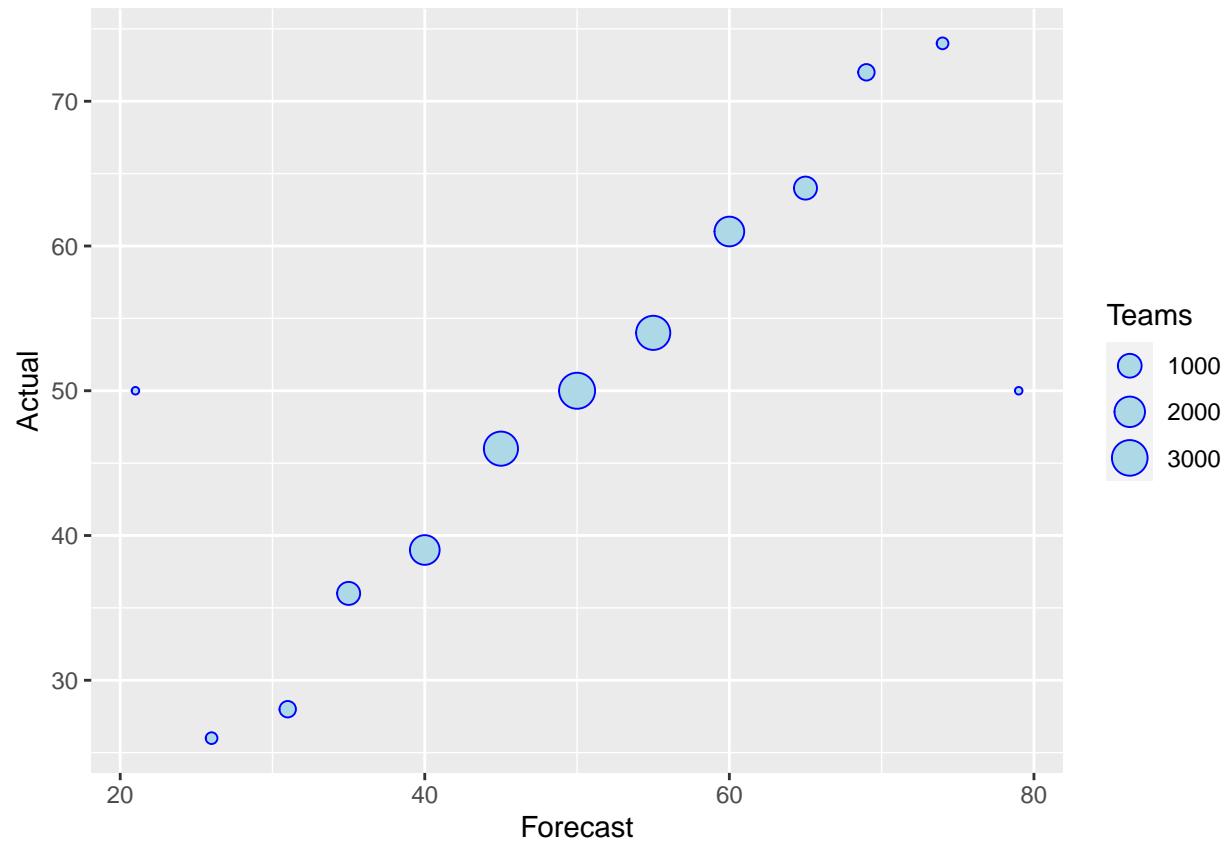
First Scatterplot

```
library(ggplot2)
ggplot(data = x) + geom_point(mapping = aes(x = Forecast, y = Actual, size = Teams))
```
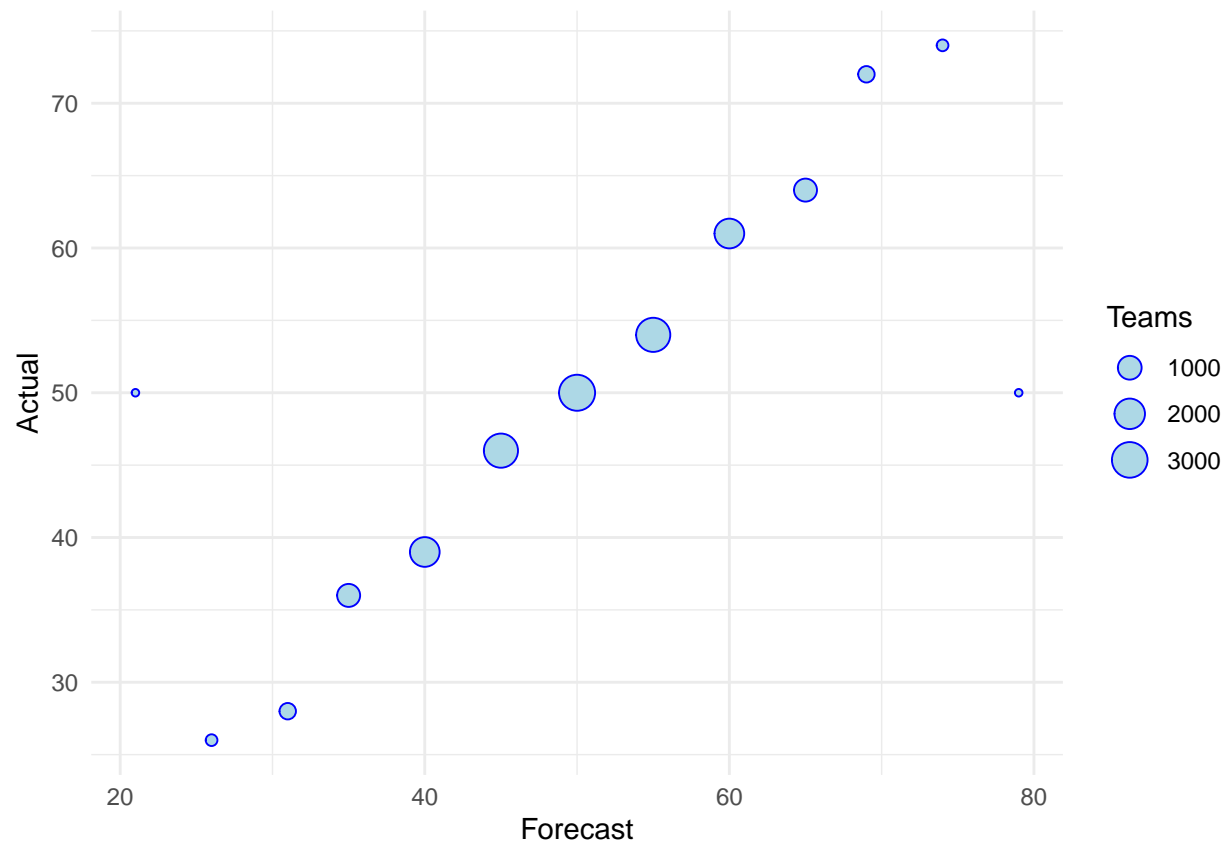
Second Scatterplot

```
calplot <- ggplot(data = x) +
  geom_point(aes(Forecast, Actual, size = Teams), color = 'blue', fill = 'lightblue', shape = 21)
print(calplot)
```
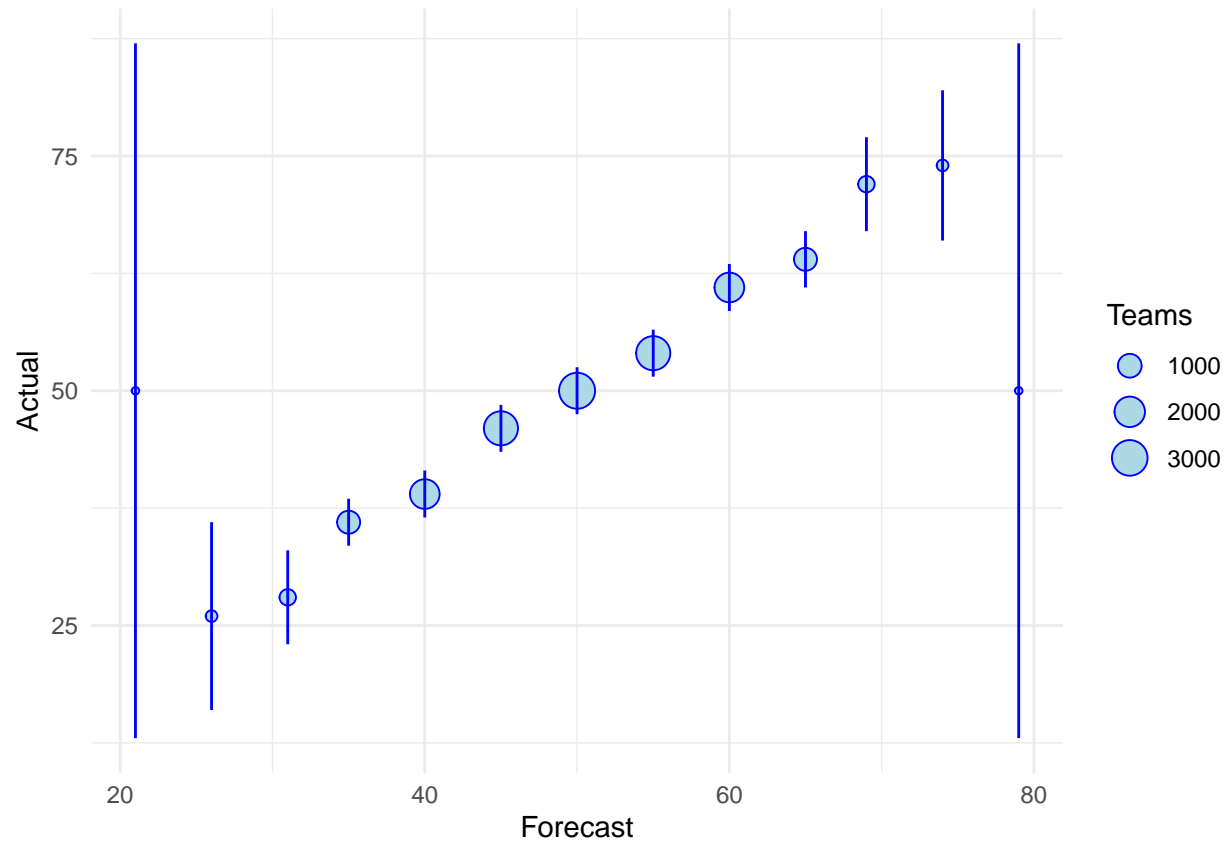
Change Theme

```
calplot <- calplot + theme_minimal()
print(calplot)
```
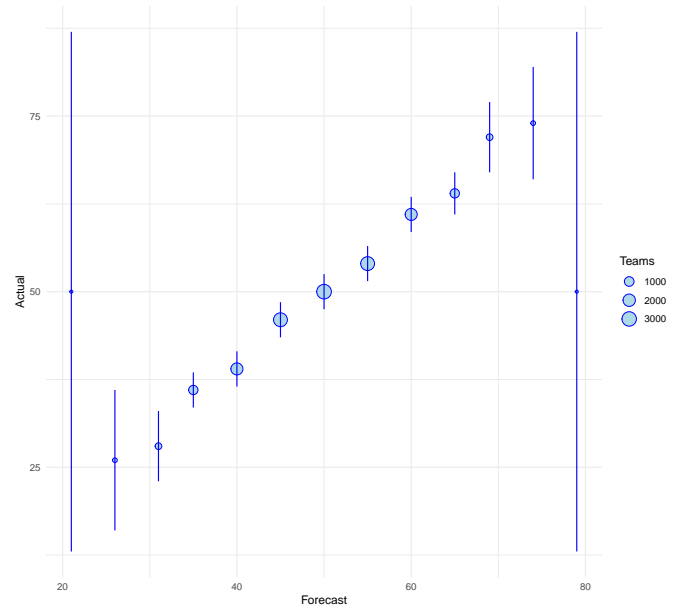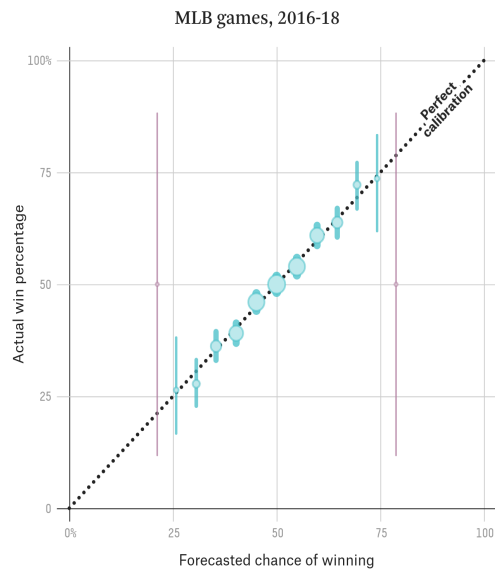
Adding the 95% CI

```
calplot <- calplot +
  geom_linerange(aes(Forecast, ymin = Actual - Interval, ymax = Actual + Interval), color = 'blue')
print(calplot)
```
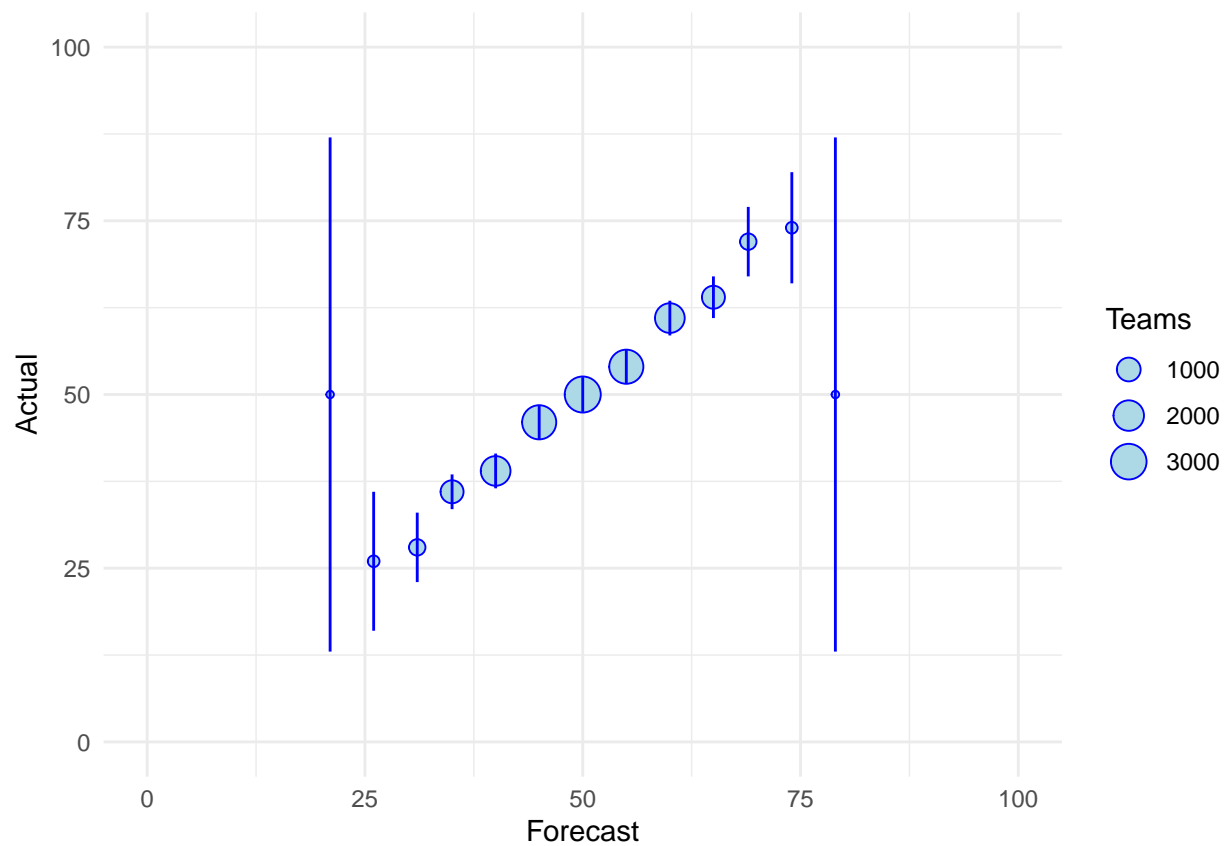
Comparision 1

```r
library(ggpubr)
library(patchwork)
library(png)

screencap <- ggplot() + background_image(readPNG('538-calibration-chart.png'))
(screencap + calplot)
```
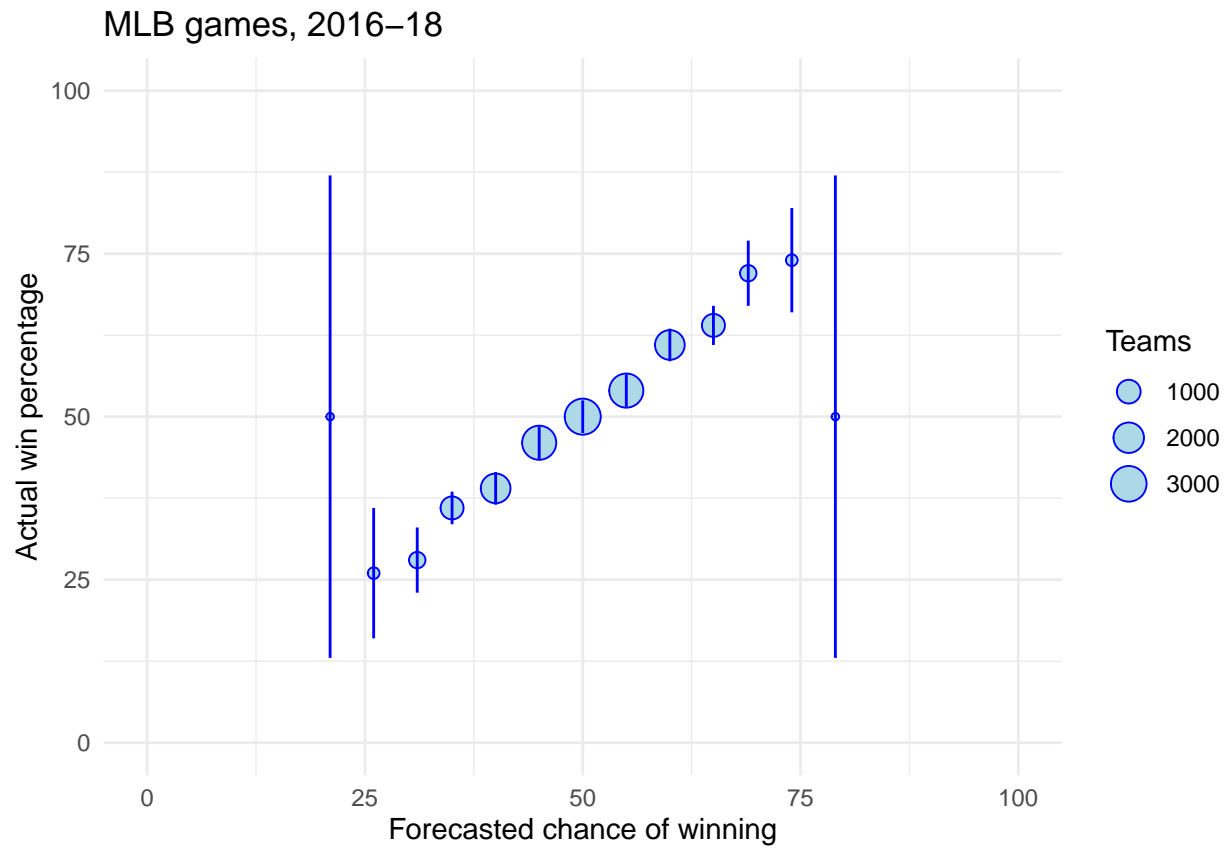
MLB games, 2016-18

Ensure that the points (0,0) and (100,100) are included in the chart:

```
calplot <- calplot +
  geom_blank(aes(x = 0, y = 0)) +
  geom_blank(aes(x = 100, y = 100))
print(calplot)
```
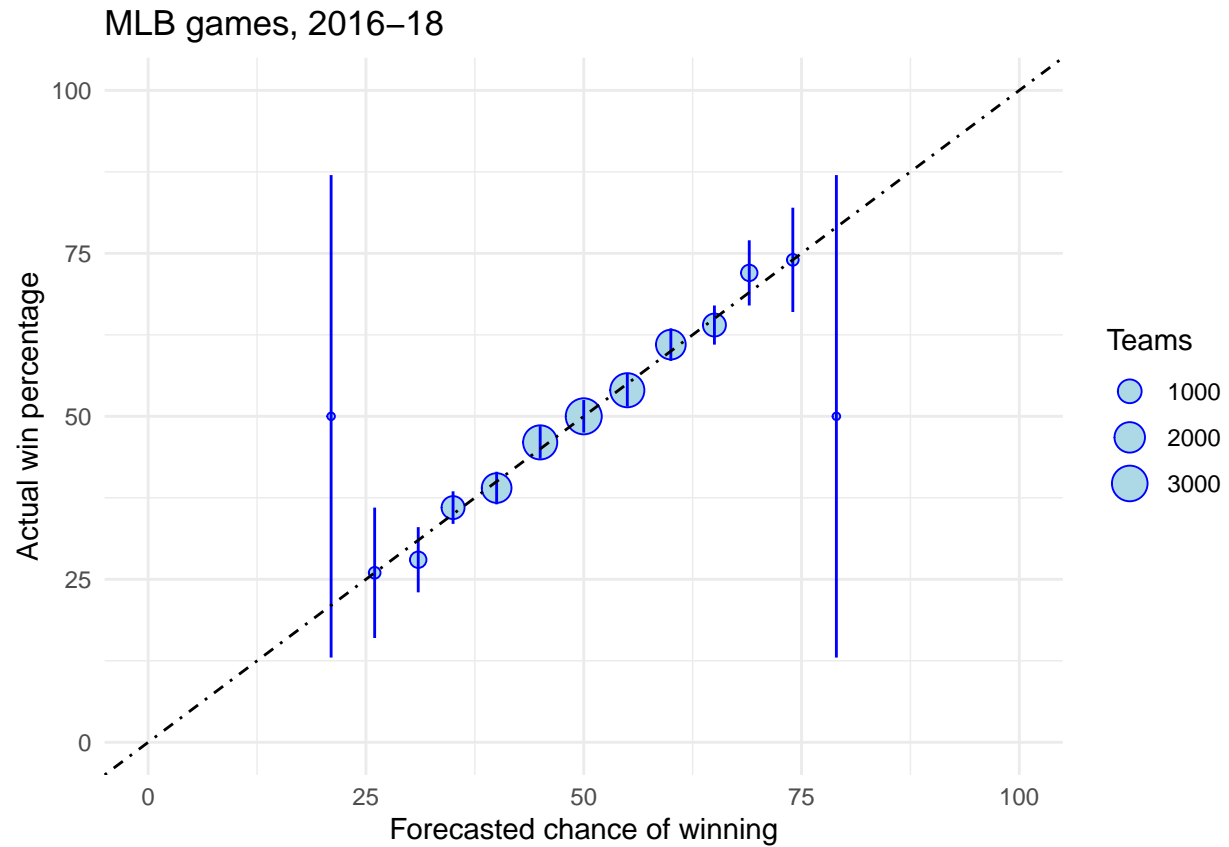
Add Labels

```
calplot <- calplot + labs(title = 'MLB games, 2016-18',
                          x = 'Forecasted chance of winning',
                          y = 'Actual win percentage')
print(calplot)
```
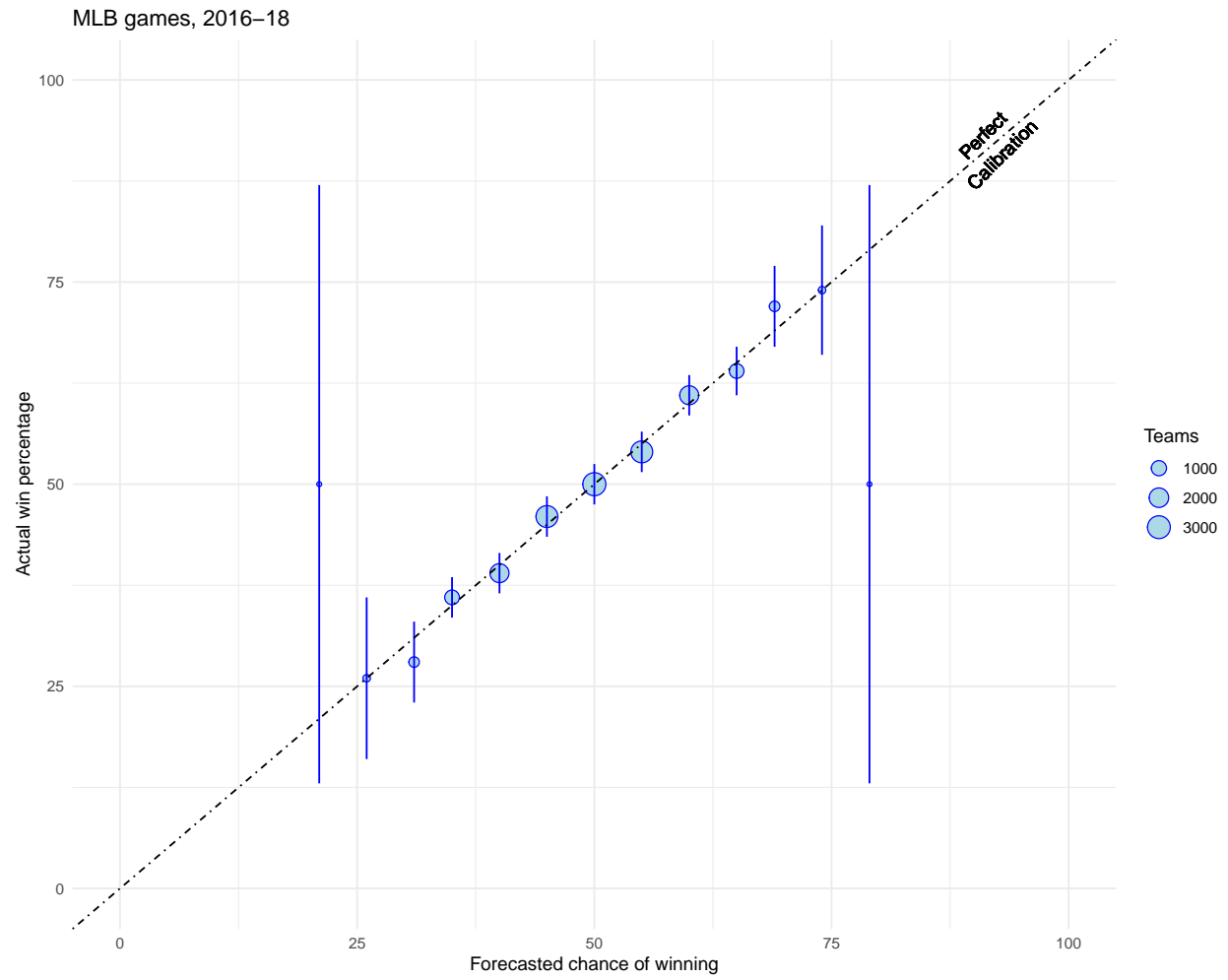

Add Perfect Calibration Line

```
calplot <- calplot + geom_abline(aes(intercept = 0, slope = 1), lty = 4)
print(calplot)
```

MLB games, 2016–18

Add Perfect Calibration Text

```
calplot <- calplot +
  geom_text(aes(x = 92, y = 92, label = 'Perfect\nCalibration', angle = 45))
print(calplot)
```

MLB games, 2016–18

Comparision 2

MLB games, 2016-18

MLB games, 2016–18