In [4]:
```
# Question 4

nfl = read.csv("NHL.csv",stringsAsFactors = F)
library(ggplot2)
library(gridExtra)

## PART 1
lin.reg = lm(A~G, data = nfl)
summary(lin.reg)
coef(lin.reg)
g1 <- ggplot(nfl, aes(x = G, y = A, color = GP, size = G)) +
geom_point() + scale_color_gradient(low = "blue", high = "red") +
geom_abline(intercept = 494.513,slope = 0.5202063)
g1


## PART 2
lin.reg1 = lm(I(A-1963) ~ I(G - 894) + 0, data = nfl)
summary(lin.reg1)
coef(lin.reg1)
g2 <- ggplot(nfl, aes(x = G, y = A, color = GP, size = G)) +
geom_point() + scale_color_gradient(low = "blue", high = "red") +
geom_abline(intercept = 0 , slope = 2.684662 )
g2

## PART 3
# Stats for patrick kane
kane = list(88,("Patrick Kane"),NA,NA,NA,NA,as.integer(728),as.integer
(282),as.integer(458),as.integer(740),as.integer(81),as.integer(282),a
s.integer(85),as.integer(0),as.integer(44),NA,as.integer(3),as.integer
(2269))
nfl = rbind(nfl,kane)

lin.reg2 = lm(I(A-458) ~ I(G - 282) + 0, data = nfl)
summary(lin.reg2)
coef(lin.reg2)
g3 <- ggplot(nfl, aes(x = G, y = A, color = GP, size = G)) +
geom_point() + scale_color_gradient(low = "blue", high = "red") +
geom_abline(intercept = 0 , slope = 1.150172 )
g3


## Part 4
lin.reg0 = lm(A~G + 0, data = nfl)
summary(lin.reg0)

lin.reg10 = lm(I(A-1963) ~ I(G - 894) + 0, data = nfl)
summary(lin.reg10)
```

```
lin.reg20 = lm(I(A-458) ~ I(G - 282) + 0, data = nfl)
summary(lin.reg20)

## Part 5
Table = data.frame(Model.Name = c("lin.reg0","lin.reg10","lin.reg20"),
RSquared = c("0.8975","0.9094","0.5787"), PValue = c("< 2.2e-16","< 2.
2e-16","< 2.2e-16"))
Table
```

```
Call:
lm(formula = A ~ G, data = nfl)

Residuals:
    Min      1Q  Median      3Q     Max
-356.51 -158.59  -10.29  125.08 1003.42

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 494.5135    74.6410   6.625 1.88e-09 ***
G             0.5202     0.1508   3.449 0.000832 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 207 on 98 degrees of freedom
Multiple R-squared:  0.1082,    Adjusted R-squared:  0.09913
F-statistic: 11.89 on 1 and 98 DF,  p-value: 0.0008317
```

(Intercept) 494.513523295988

G 0.520206294863515

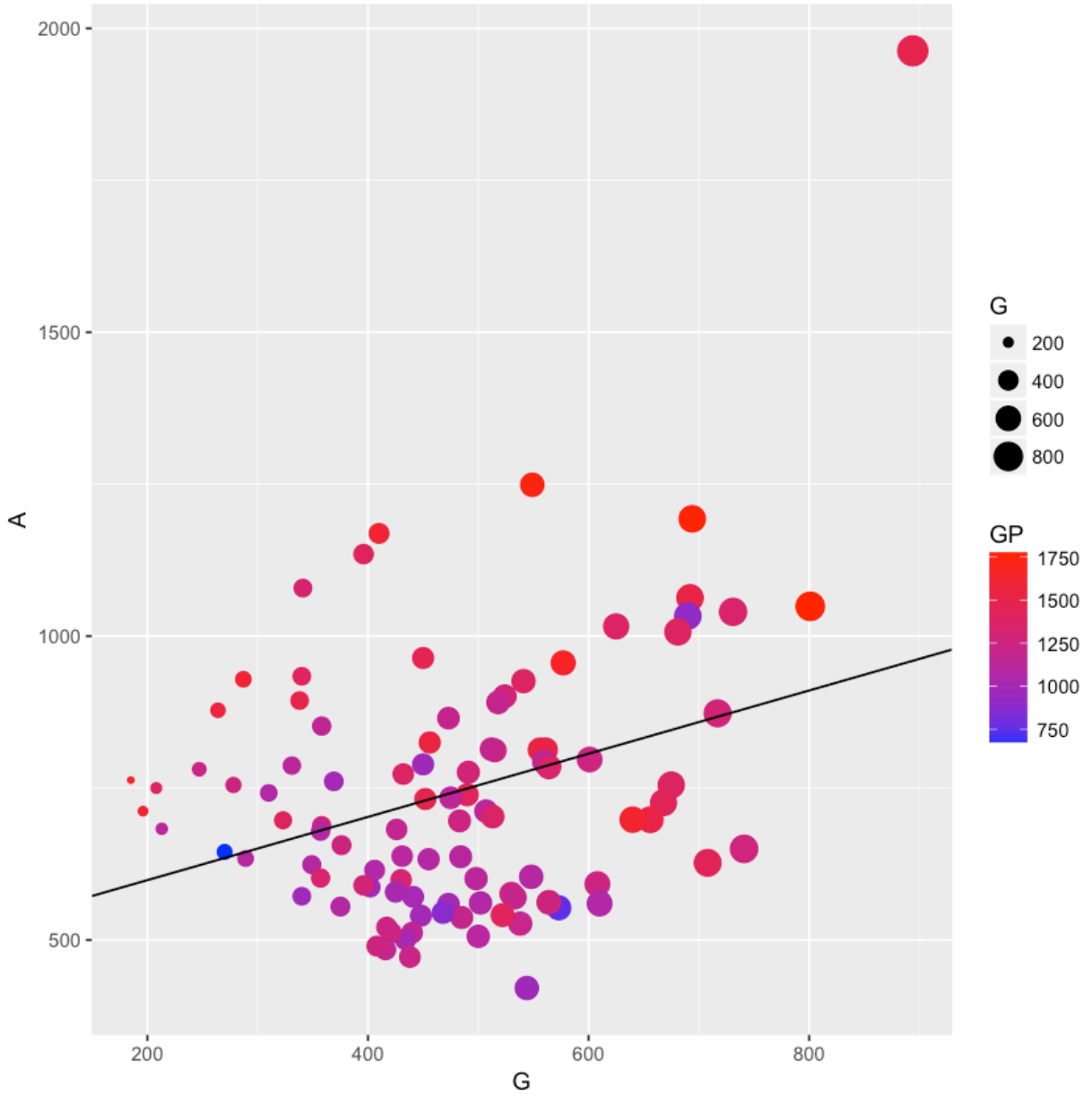
```
Call:
lm(formula = I(A - 1963) ~ I(G - 894) + 0, data = nfl)

Residuals:
   Min     1Q Median     3Q    Max
-902.2 -333.4 -145.4  103.3  703.4

Coefficients:
           Estimate Std. Error t value Pr(>|t|)
I(G - 894)   2.6847     0.0853   31.48   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 375.7 on 99 degrees of freedom
Multiple R-squared:  0.9091,    Adjusted R-squared:  0.9082
F-statistic: 990.7 on 1 and 99 DF,  p-value: < 2.2e-16
```

I(G - 894): 2.68466238522409


2000
1500
1000
500
A
200
400
600
800
G
G
200
400
600
800
GP
1750
1500
1250
1000
750

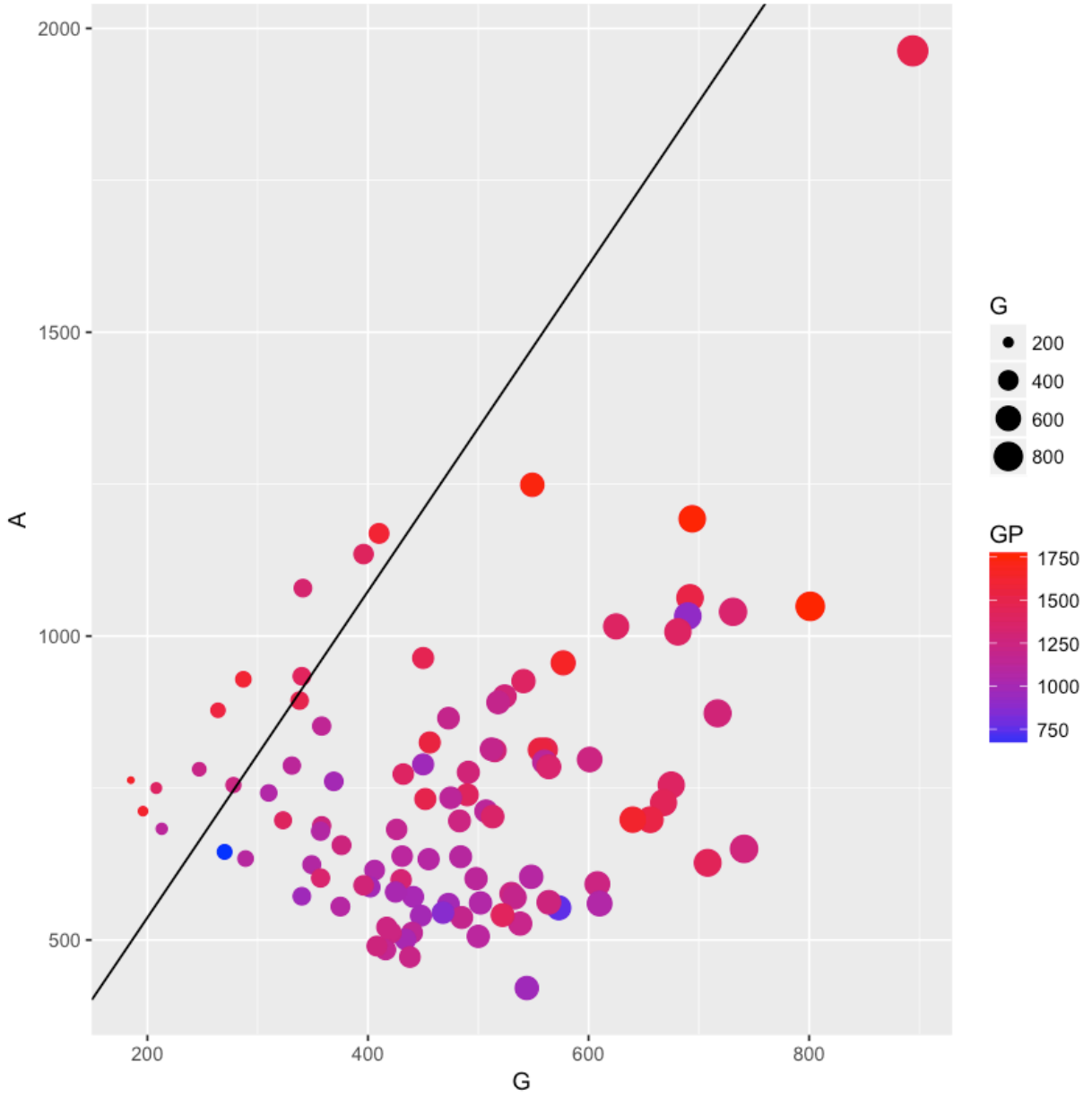
```
Call:
lm(formula = I(A - 458) ~ I(G - 282) + 0, data = nfl)

Residuals:
    Min      1Q  Median      3Q     Max
-338.35 -112.92   39.85  167.95  801.09

Coefficients:
           Estimate Std. Error t value Pr(>|t|)
I(G - 282)   1.1502     0.0973   11.82   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 230.8 on 100 degrees of freedom
Multiple R-squared:  0.5829,    Adjusted R-squared:  0.5787
F-statistic: 139.7 on 1 and 100 DF,  p-value: < 2.2e-16
```

I(G - 282): 1.15017224294842


2000
1500
1000
500
A
200
400
600
800
G
G
200
400
600
800
GP
1750
1500
1250
1000
750

```
Call:
lm(formula = A ~ G + 0, data = nfl)

Residuals:
   Min     1Q Median     3Q    Max
-447.3 -137.1    3.6  157.9  639.2

Coefficients:
  Estimate Std. Error t value Pr(>|t|)
G  1.48080    0.04977   29.75   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 246.7 on 100 degrees of freedom
Multiple R-squared:  0.8985,    Adjusted R-squared:  0.8975
F-statistic: 885.4 on 1 and 100 DF,  p-value: < 2.2e-16

Call:
lm(formula = I(A - 1963) ~ I(G - 894) + 0, data = nfl)

Residuals:
   Min     1Q Median     3Q    Max
-902.9 -329.7 -141.1  121.8  700.4

Coefficients:
           Estimate Std. Error t value Pr(>|t|)
I(G - 894)  2.68039    0.08412   31.86   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 374.1 on 100 degrees of freedom
Multiple R-squared:  0.9103,    Adjusted R-squared:  0.9094
F-statistic:  1015 on 1 and 100 DF,  p-value: < 2.2e-16

Call:
lm(formula = I(A - 458) ~ I(G - 282) + 0, data = nfl)

Residuals:
    Min      1Q  Median      3Q     Max
-338.35 -112.92   39.85  167.95  801.09

Coefficients:
           Estimate Std. Error t value Pr(>|t|)
I(G - 282)   1.1502     0.0973   11.82   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 230.8 on 100 degrees of freedom
Multiple R-squared:  0.5829,    Adjusted R-squared:  0.5787
F-statistic: 139.7 on 1 and 100 DF,  p-value: < 2.2e-16
```

Model.Name	RSquared	PValue
lin.reg0	0.8975	< 2.2e-16
lin.reg10	0.9094	< 2.2e-16
lin.reg20	0.5787	< 2.2e-16