# R Coding Demonstration Week 5: Predicting Presidential Outcomes

Matthew Blackwell

Gov 51 (Harvard)

# Introduction

- Long tradition of presidential election forecasting in political science
- Traditionally, “fundamentals”-based (measured by June of election year).
- “Time-for-change” model by Alan Abramowitz is very simple function of three predictors:
    - Party of incumbent president.
    - Popularity of the incumbent president
    - Economic conditions early in the election year.
- We usually predict incumbent-party share of the two-party vote.
    - Don’t have to worry as much about third parties.

# Data

- Load the data

```
pres <- read.csv("data/pres_data.csv")
```

- Variables:

| Variable Name | Description |
|---|---|
| `year` | Election year |
| `deminc` | 1=incumbent is a Democrat |
| `incvote` | Percent of the two-party vote for the incumbent party |
| `q2gdp` | Second-quarter change in real GDP in election year |
| `juneapp` | Net approval of incumbent president in June of election year |

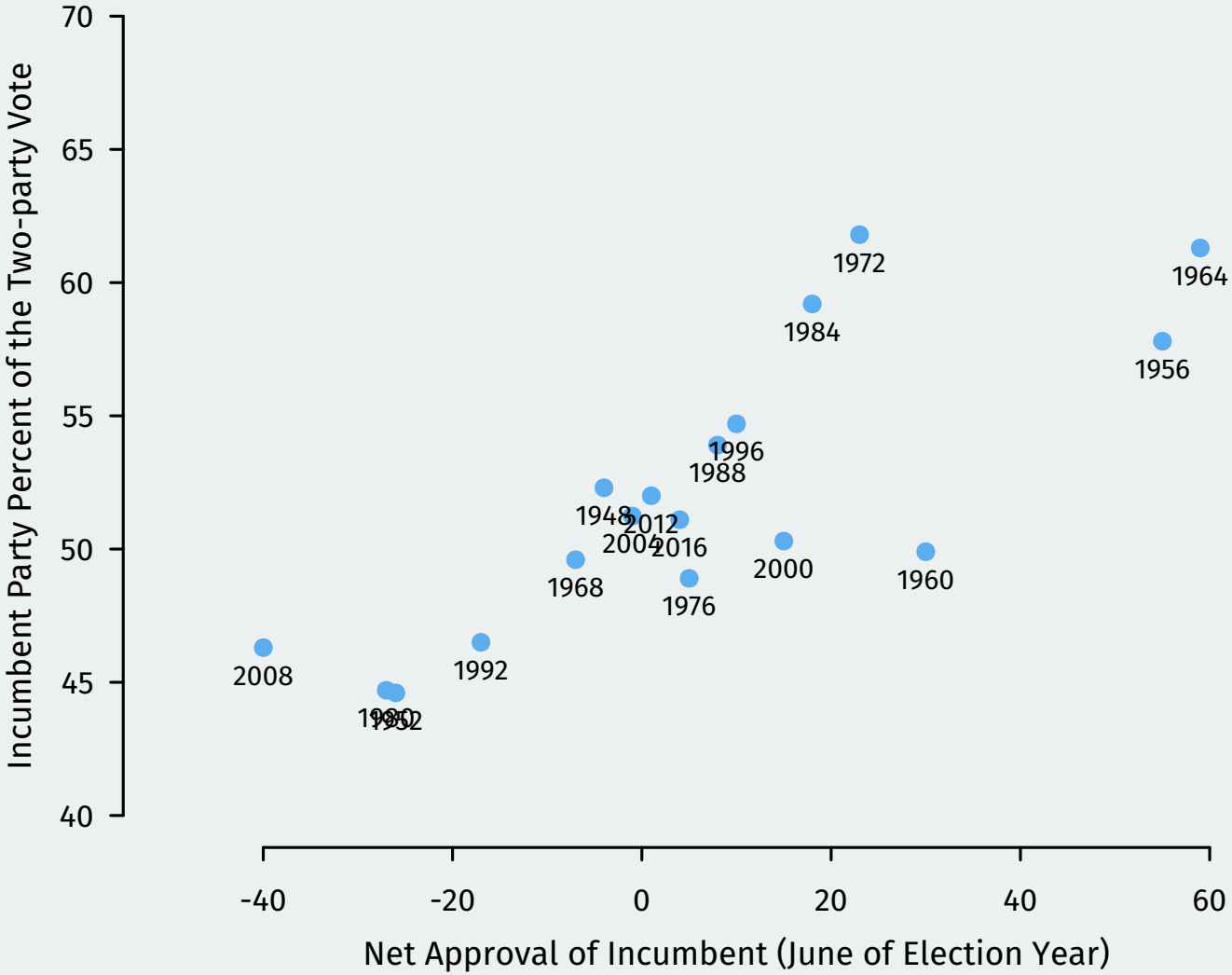
# Question 1

Create a scatter plot of vote for the incumbent party on the y-axis and June net approval on the x-axis. Add labels for the years to each point.

# Answer 1 (cont'd)

```
plot(x = pres$juneapp, y = pres$incvote, pch = 19,
     col = "steelblue2", frame = FALSE, las = 1,
     xlab = "Net Approval of Incumbent (June of Election Year)",
     ylab = "Incumbent Party Percent of the Two-party Vote",
     xlim = c(-50, 70), ylim = c(40, 70))
text(x = pres$juneapp, y = pres$incvote, labels = pres$year,
     pos = 1, cex = 0.8)
```

# Answer 1 (cont'd)



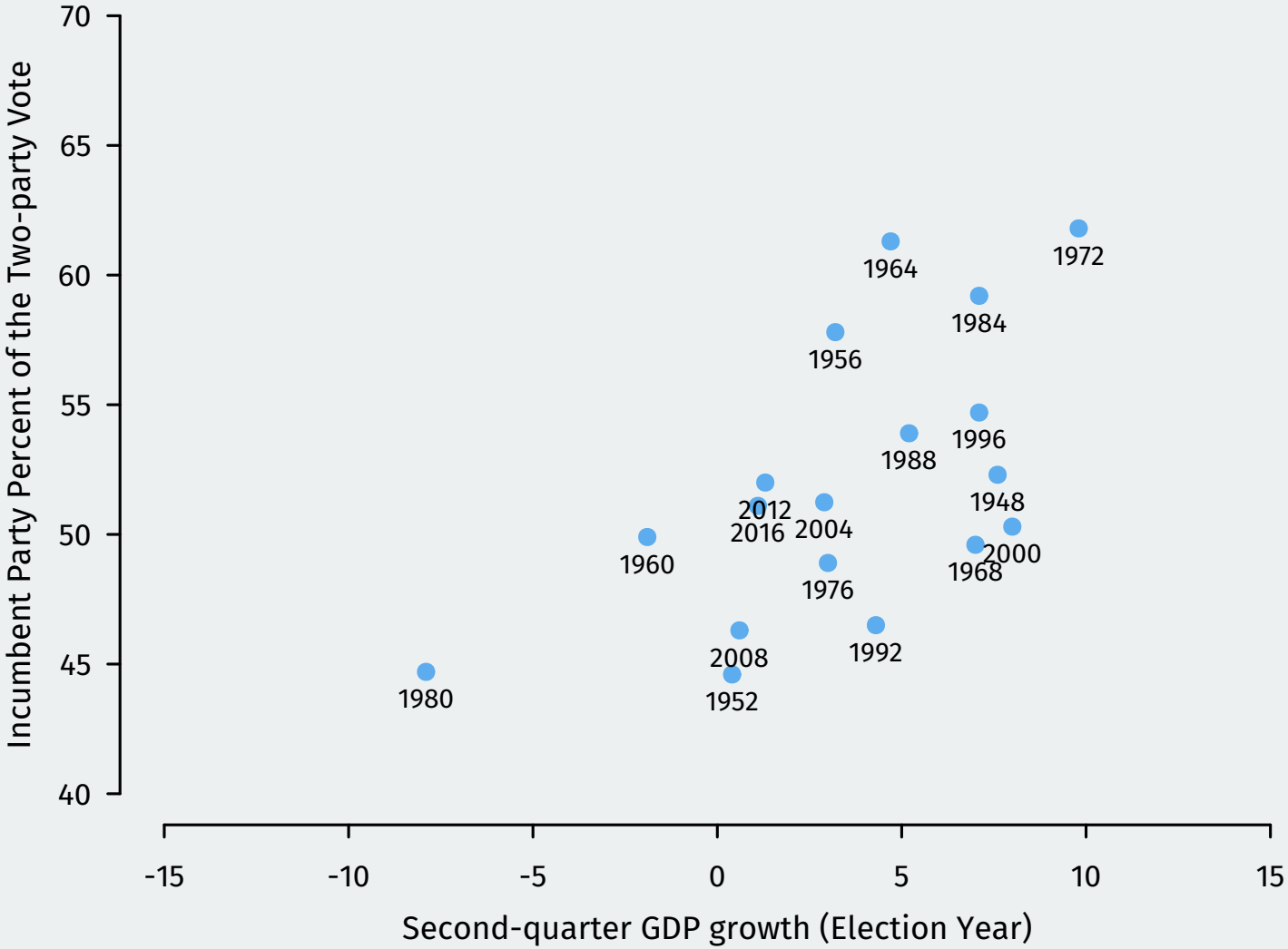
# Question 2

Create a scatter plot of vote for the incumbent party on the y-axis and second-quarter GDP growth on the x-axis. Add labels for the years to each point.

# Answer 2

```
plot(x = pres$q2gdp, y = pres$incvote, pch = 19,
     col = "steelblue2", frame = FALSE, las = 1,
     xlab = "Second-quarter GDP growth (Election Year)",
     ylab = "Incumbent Party Percent of the Two-party Vote",
     xlim = c(-15, 15), ylim = c(40, 70))
text(x = pres$q2gdp, y = pres$incvote, labels = pres$year,
     pos = 1, cex = 0.8)
```

# Answer 2 (cont'd)



# Question 3

Run a regression with incumbent vote as the dependent variable and June net approval as the independent variable. Interpret the estimated coefficient on net approval.

Add the line of best fit from this regression onto the previous scatterplot.

# Answer 3

```
app_fit <- lm(incvote ~ juneapp, data = pres)
app_fit
```

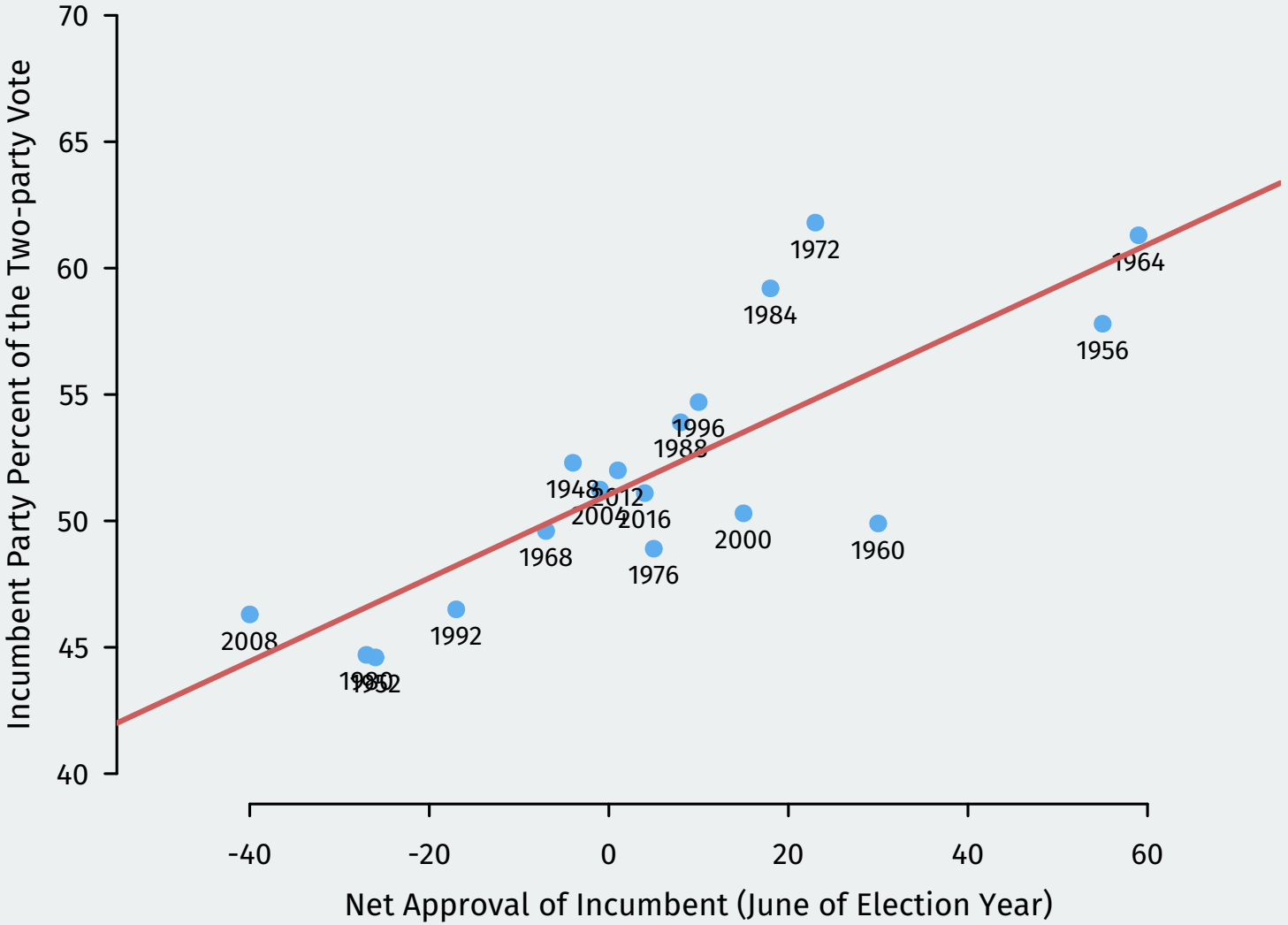
```
##
## Call:
## lm(formula = incvote ~ juneapp, data = pres)
##
## Coefficients:
## (Intercept)      juneapp
##      51.037        0.165
```

For every 1 percentage point increase in net approval, we see a 0.16 increase in the vote share for the incumbent party.

# Answer 3 (cont'd)

```
plot(x = pres$juneapp, y = pres$incvote, pch = 19,
     col = "steelblue2", frame = FALSE, las = 1,
     xlab = "Net Approval of Incumbent (June of Election Year)",
     ylab = "Incumbent Party Percent of the Two-party Vote",
     xlim = c(-50, 70), ylim = c(40, 70))
text(x = pres$juneapp, y = pres$incvote, labels = pres$year,
     pos = 1, cex = 0.8)
abline(app_fit, col = "indianred", lwd = 2)
```

# Answer 3 (cont'd)



# Question 4

Run a regression with incumbent vote as the dependent variable and second-quarter GDP growth as the independent variable. Interpret the estimated coefficient on net approval.

Add the line of best fit from this regression onto the previous scatterplot.

# Answer 4

```
gdp_fit <- lm(incvote ~ q2gdp, data = pres)
gdp_fit
```

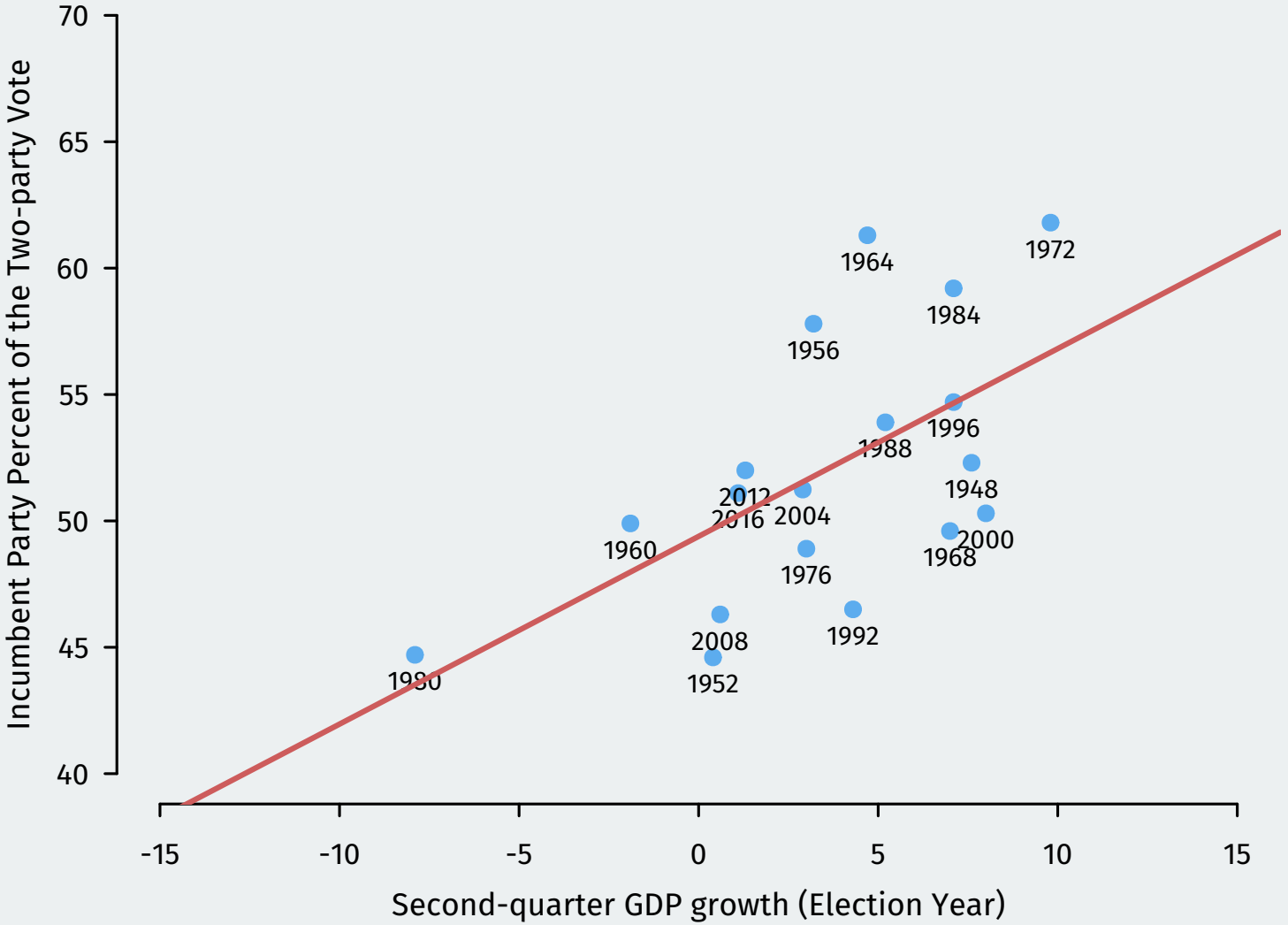
```
##
## Call:
## lm(formula = incvote ~ q2gdp, data = pres)
##
## Coefficients:
## (Intercept)        q2gdp
##      49.388        0.743
```

For every 1 percentage point increase in second-quarter GDP growth, we see a 0.74 increase in the vote share for the incumbent party.

# Answer 4 (cont'd)

```
plot(x = pres$q2gdp, y = pres$incvote, pch = 19,
     col = "steelblue2", frame = FALSE, las = 1,
     xlab = "Second-quarter GDP growth (Election Year)",
     ylab = "Incumbent Party Percent of the Two-party Vote",
     xlim = c(-15, 15), ylim = c(40, 70))
text(x = pres$q2gdp, y = pres$incvote, labels = pres$year,
     pos = 1, cex = 0.8)
abline(gdp_fit, col = "indianred", lwd = 2)
```

# Answer 4 (cont'd)



# Question 5

Use the coefficients from the regression of incumbent vote on net approval to get a predicted value for Trump's percent of the two-party vote in 2020.

# Answer 5

```
juneapp_2020 <- pres$juneapp[pres$year == 2020]

coef(app_fit)["(Intercept)"] +
  juneapp_2020 * coef(app_fit)["juneapp"]
```

```
## (Intercept)
##        48.1
```

# Question 6

Use the coefficients from the regression of incumbent vote on second-quarter GDP growth to get a predicted value for Trump's percent of the two-party vote in 2020. Compare it to the prediction from the last model. Which do you think is a better prediction?

# Answer 6

```
q2gdp_2020 <- pres$q2gdp[pres$year == 2020]

coef(gdp_fit)["(Intercept)"] +
  q2gdp_2020 * coef(gdp_fit)["q2gdp"]
```

```
## (Intercept)
##        26.1
```

# Question 7

Calculate the root mean squared error of the regression of incumbent part vote on incumbent net approval.

# Answer 7

```
app_resids <- residuals(app_fit)
app_rmse <- sqrt(mean(app_resids ^ 2))
app_rmse
```

```
## [1] 3.01
```

# Question 8

Calculate the root mean squared error of the regression of incumbent part vote on second-quarter GDP growth. Which variable does a better job predicting the outcome?

# Answer 8

```
gdp_resids <- residuals(gdp_fit)
gdp_rmse <- sqrt(mean(gdp_resids ^ 2))
gdp_rmse
```

```
## [1] 4.11
```

# Question 9 (Challenge question)

Perform the two regressions from above on all elections before 2016 (that is, omit 2016 from these regressions). For each regression, calculate the predicted 2016 incumbent vote share. Which model performs better?

# Answer 9

```
pres_no16 <- subset(pres, year < 2016)
app_fit_no16 <- lm(incvote ~ juneapp, data = pres_no16)
gdp_fit_no16 <- lm(incvote ~ q2gdp, data = pres_no16)

juneapp_2016 <- pres$juneapp[pres$year == 2016]
q2gdp_2016 <- pres$q2gdp[pres$year == 2016]
incvote_2016 <- pres$incvote[pres$year == 2016]

app_pred_2016 <- coef(app_fit_no16)[1] +
  juneapp_2016 * coef(app_fit_no16)[2]

gdp_pred_2016 <- coef(gdp_fit_no16)[1] +
  q2gdp_2016 * coef(gdp_fit_no16)[2]

incvote_2016 - app_pred_2016
```

```
## (Intercept)
##      -0.632
```

```
incvote_2016 - gdp_pred_2016
```

```
## (Intercept)
##       0.968
```