Day4 final exercise solutions

Michael Levy

September 17, 2015

- Download the quarterly presidential approval ratings dataset from SmartSite. Load and inspect it, making any necessary changes.

```
library(tidyr)
library(dplyr)
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 3.2.2
```

```
library(stargazer)
setwd('~/Dropbox/Teaching/RBootcamp/Day4/')
d = readRDS('data/pres.RDS')
head(d)
```

```
##         president Q1 Q2 Q3 Q4
## 1945       Truman NA NA 71 38
## 1946       Truman 87 32 62 46
## 1947       Truman 82 59 61 36
## 1948       Truman 75 74 57 49
## 1949       Truman 63 75 72 35
## 1950       Truman 50 60 83 44
```

```
d$year = rownames(d)
```

- Is it tidy? If not, tidy it.

```
d = gather(d, quarter, approval, Q1:Q4)
head(d)
```

```
##    president year quarter approval
## 1     Truman 1945      Q1       NA
## 2     Truman 1946      Q1       87
## 3     Truman 1947      Q1       82
## 4     Truman 1948      Q1       75
## 5     Truman 1949      Q1       63
## 6     Truman 1950      Q1       50
```

- Graphically explore these questions to develop a testable hypothesis:
    - Do the presidents differ in their approval ratings?

```
ggplot(d,
       aes(x = reorder(president, approval, median, na.rm = TRUE),
           y = approval)) +
    geom_boxplot()
```

```
## Warning in loop_apply(n, do.ply): Removed 6 rows containing non-finite
## values (stat_boxplot).
```
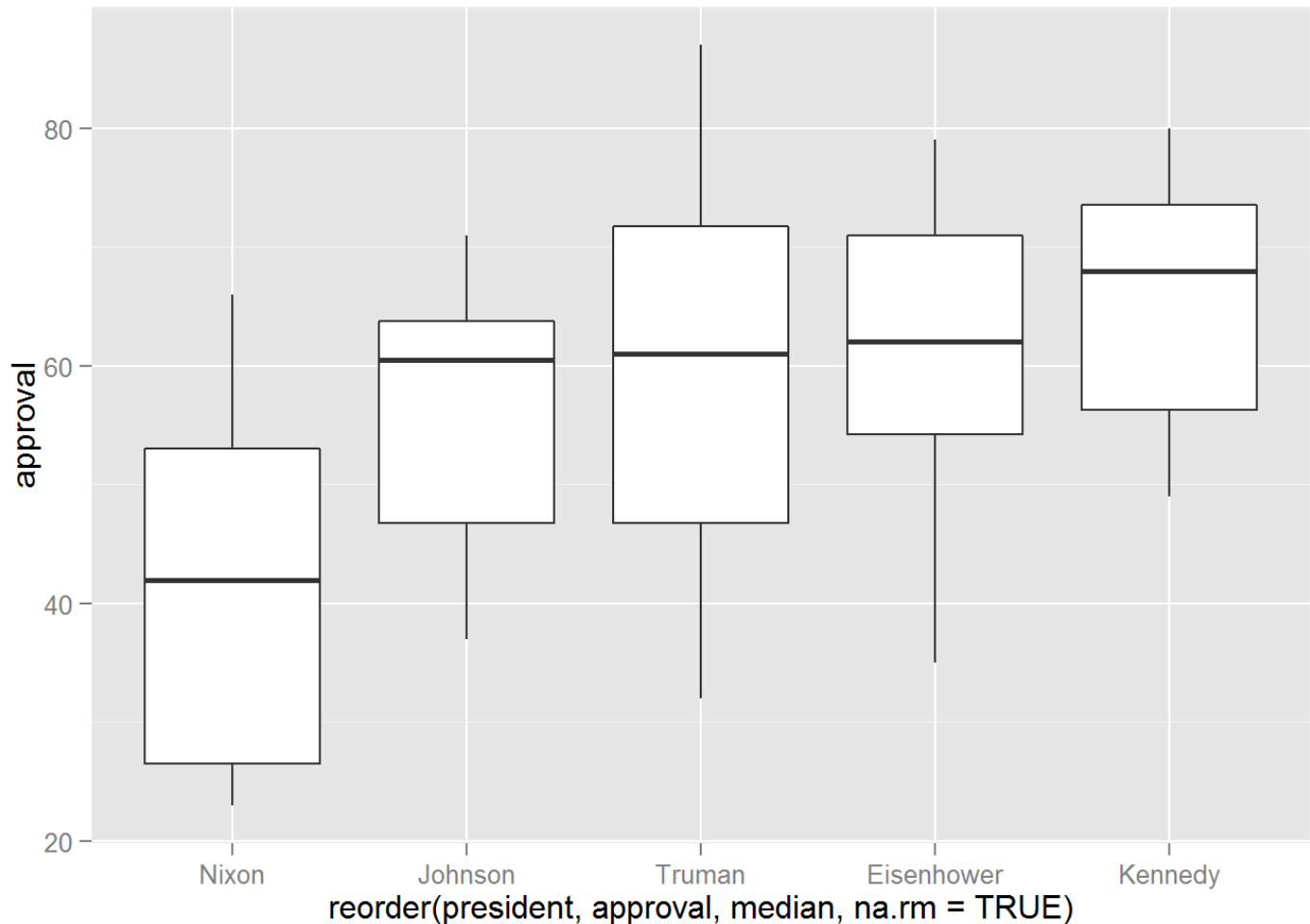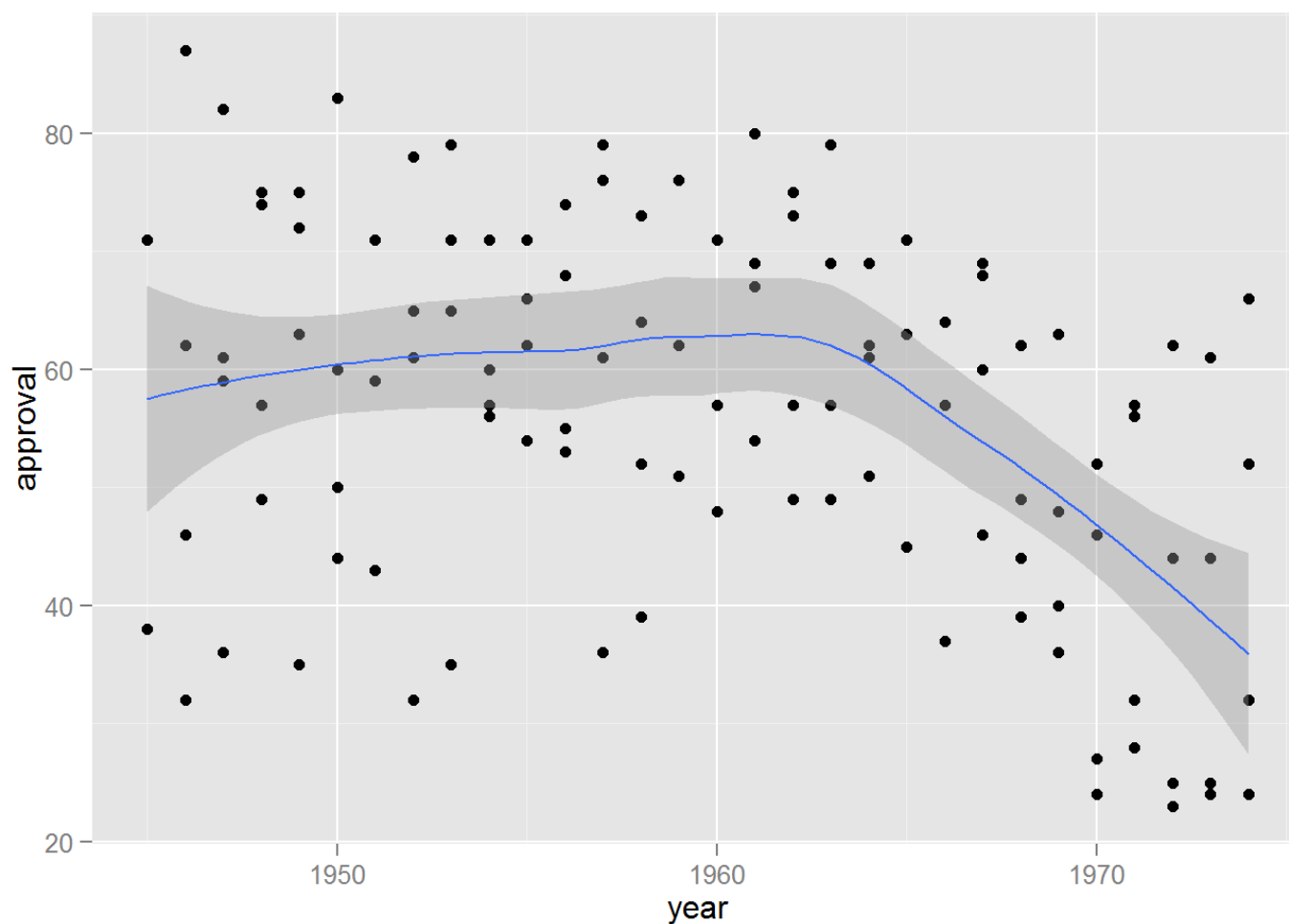

Sure looks like it. At least one does!

```
+ Do approval ratings vary over time?
```

```
d$year = as.integer(d$year)
ggplot(d, aes(x = year, y = approval)) +
    geom_point() +
    geom_smooth()
```

```
## Warning in loop_apply(n, do.ply): Removed 6 rows containing missing values
## (stat_smooth).
```

```
## Warning in loop_apply(n, do.ply): Removed 6 rows containing missing values
## (geom_point).
```
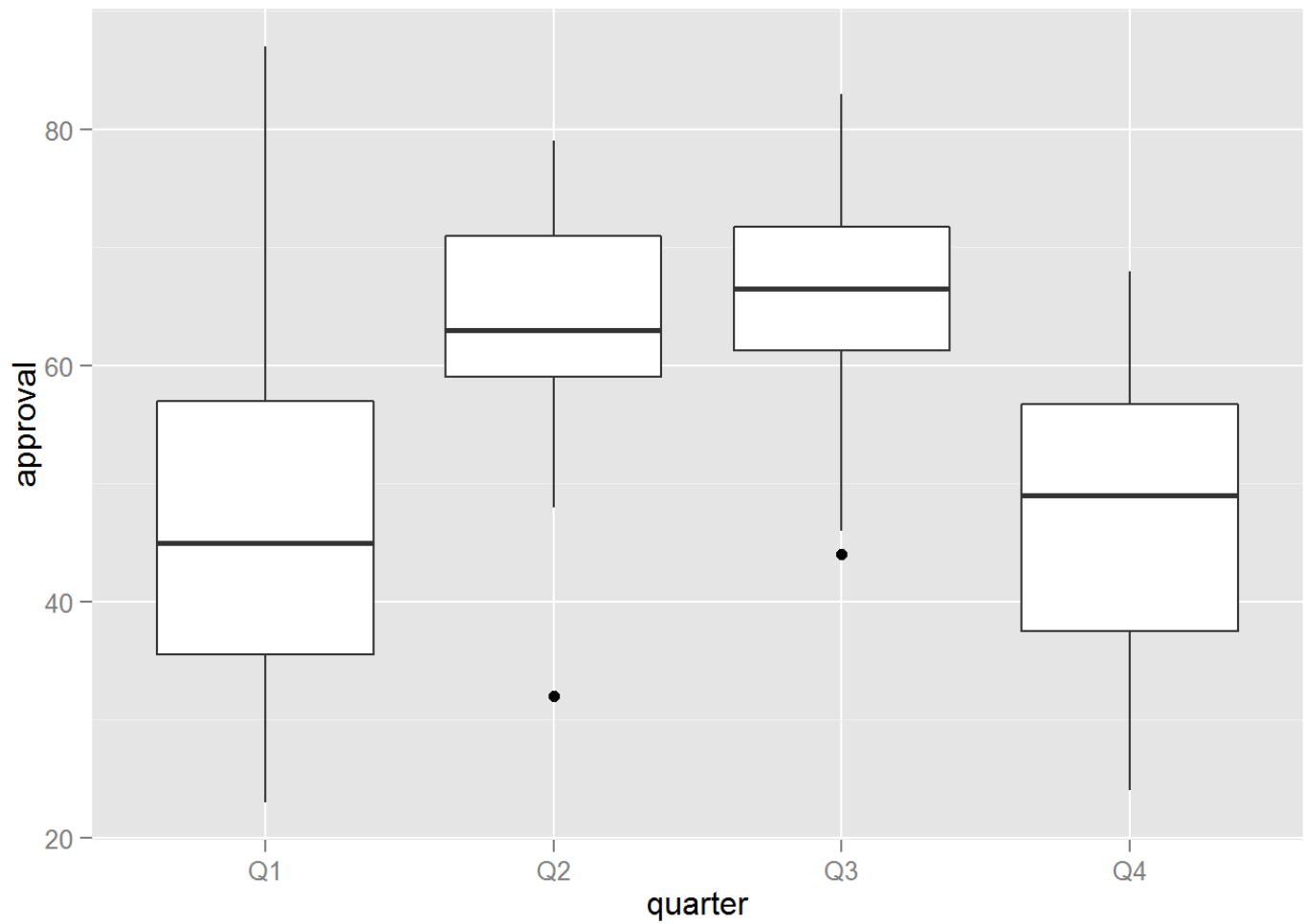

Hmm… something happens in the mid-60s after which approval ratings fall off.

```
+ Is there a seasonal effect of quarter on approval ratings?
```

```
ggplot(d, aes(x = quarter, y = approval)) +
    geom_boxplot()
```

```
## Warning in loop_apply(n, do.ply): Removed 6 rows containing non-finite
## values (stat_boxplot).
```

Looks like people get grumpy in the winter and give lower approval ratings.

- Test your hypothesis using a linear model.

```
m = lm(approval ~ quarter, d)
summary(m)
```

```
##
## Call:
## lm(formula = approval ~ quarter, data = d)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -32.000  -8.517  -0.018   8.350  39.444
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  47.5556     2.5217  18.858  < 2e-16 ***
## quarterQ2    16.4444     3.5043   4.693 7.81e-06 ***
## quarterQ3    17.8444     3.4760   5.134 1.24e-06 ***
## quarterQ4    -0.5198     3.5343  -0.147    0.883
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 13.1 on 110 degrees of freedom
##   (6 observations deleted due to missingness)
## Multiple R-squared:  0.3147, Adjusted R-squared:  0.296
## F-statistic: 16.83 on 3 and 110 DF,  p-value: 4.546e-09
```

There is a massive effect of seasonality on approval ratings!

- Summarize your findings with one table and one plot.

```
# Note that this code chunk has the option `results = 'asis'`.
stargazer(m, type = 'html', title  = 'Linear model of approval and seasonal quarte
rs. Q1 (Jan. - Mar.) is the excuded class.')
```

Linear model of approval and seasonal quarters. Q1 (Jan. - Mar.) is the excuded class.

	Dependent variable:
	approval
quarterQ2	16.444***
	(3.504)
quarterQ3	17.844***
	(3.476)
quarterQ4	-0.520
	(3.534)
Constant	47.556***
	(2.522)
Observations	114
R^2	0.315
Adjusted R^2	0.296
Residual Std. Error	13.103 (df = 110)

F Statistic	16.835[***] (df = 3; 110)

```
meanAp = mean(d$approval, na.rm = TRUE)
ggplot(d, aes(x = quarter, y = approval)) +
    geom_boxplot() +
    geom_hline(aes(yintercept = meanAp),
               color = 'red',
               linetype = 'dashed',
               size = 1) +
    geom_text(x = 4.45, y = meanAp, angle = 90,
              hjust = -.1,
              label = "Mean approval",
              color = 'red') +
    ylim(c(0, 100)) +
    theme_bw(base_size = 18)
```

```
## Warning in loop_apply(n, do.ply): Removed 6 rows containing non-finite
## values (stat_boxplot).
```