PLSC 503 – Spring 2020
Dichotomous Covariates

February 11, 2020

"Dummies"...

- ..."naturally" dichotomous, including
  - Structural breaks
  - Proper nouns
- "Factors":

$$\texttt{partyid} = \begin{cases} 0 = \text{Labor} \\ 1 = \text{Liberal} \\ 2 = \text{Conservative} \end{cases}$$

- Ordinal variables...
- Continuous variables...

Coding Dummies

"Dummy coding":

$$\texttt{female} = \begin{cases} 0 \text{ if male} \\ 1 \text{ if female} \end{cases}$$

vs. "Effect coding":

$$\texttt{female} = \begin{cases} -1 \text{ (or } -0.5) \text{ if male} \\ 1 \text{ (or } 0.5) \text{ if female} \end{cases}$$

TL;DR: Use the former.

Dichotomous Xs

For

$$Y_i = \beta_0 + \beta_1 D_i + u_i$$

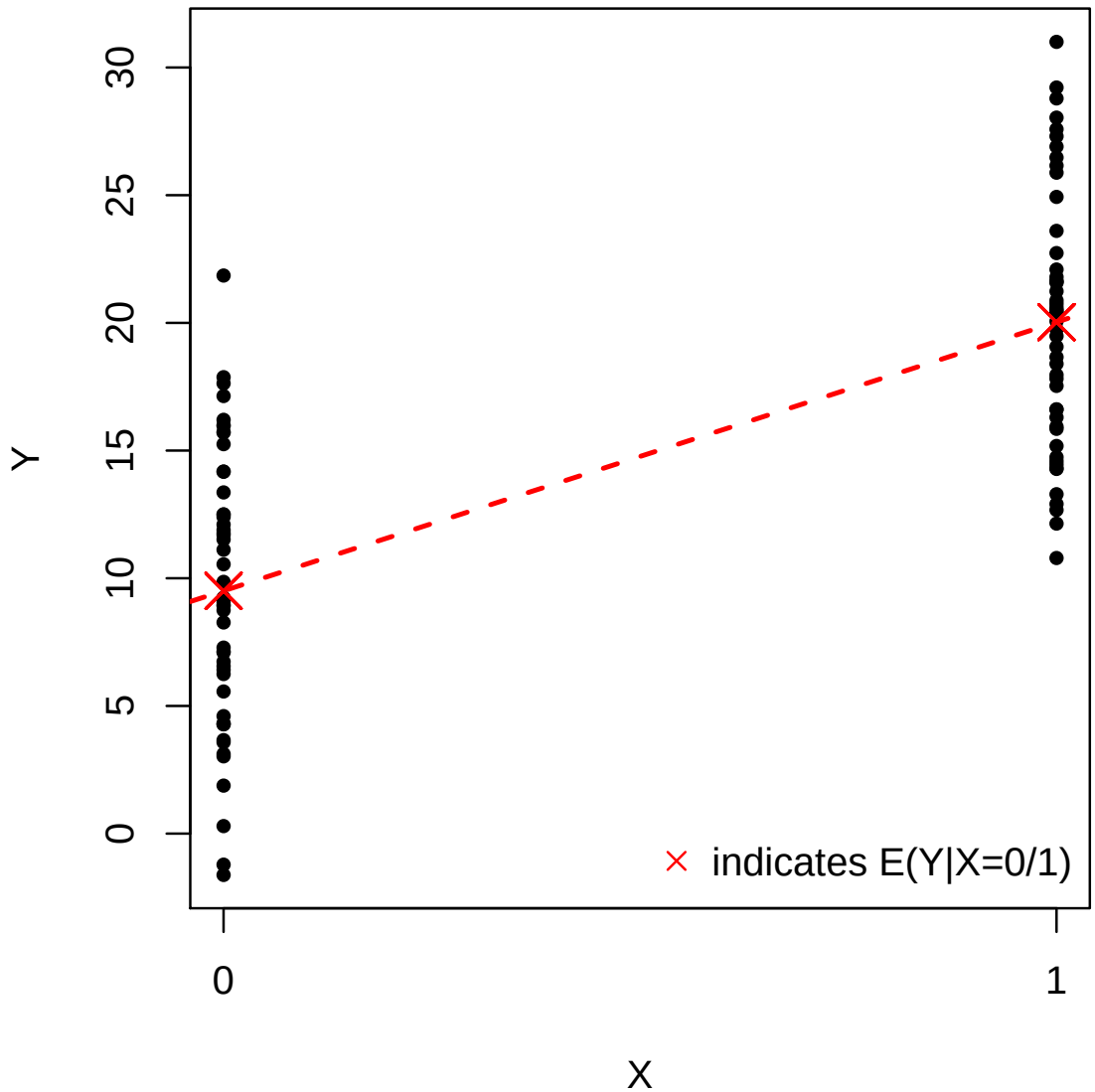
we have

$$\mathrm{E}(Y|D = 0) = \beta_0$$

and

$$\mathrm{E}(Y|D = 1) = \beta_0 + \beta_1.$$

Dichotomous X, Graphically



Many Dummies

For

$$Y_i = \beta_0 + \beta_1 D_{1i} + \beta_2 D_{2i} + ... + \beta_\ell D_{\ell i} + u_i$$

- $\mathsf{E}(Y|D_k = 0) \forall\, k \in \ell = \beta_0$,
- Otherwise, $\mathsf{E}(Y) = \beta_0 + \sum_{k=1}^{\ell} \beta_k \forall\, k\, s.t.\, D_k = 1$.

Note: where the D_ℓ are mutually exclusive and exhaustive:

- The expected values are the same as the within-group means.
- Identification requires that we either
  - omit a "reference category," or
  - omit β_0.

Dummies and Ordinal Xs

Suppose we have:

$$\texttt{gopscale} = \begin{cases} -2 = \text{Strong Democrat} \\ -1 = \text{Weak Democrat} \\ 0 = \text{Independent} \\ 1 = \text{Weak Republican} \\ 2 = \text{Strong Republican} \end{cases}$$

Might estimate:

$$\texttt{closeness}_i = 46.0 + 17.5(\texttt{gopscale}_i) + u_i$$

Dummies and Ordinal Xs

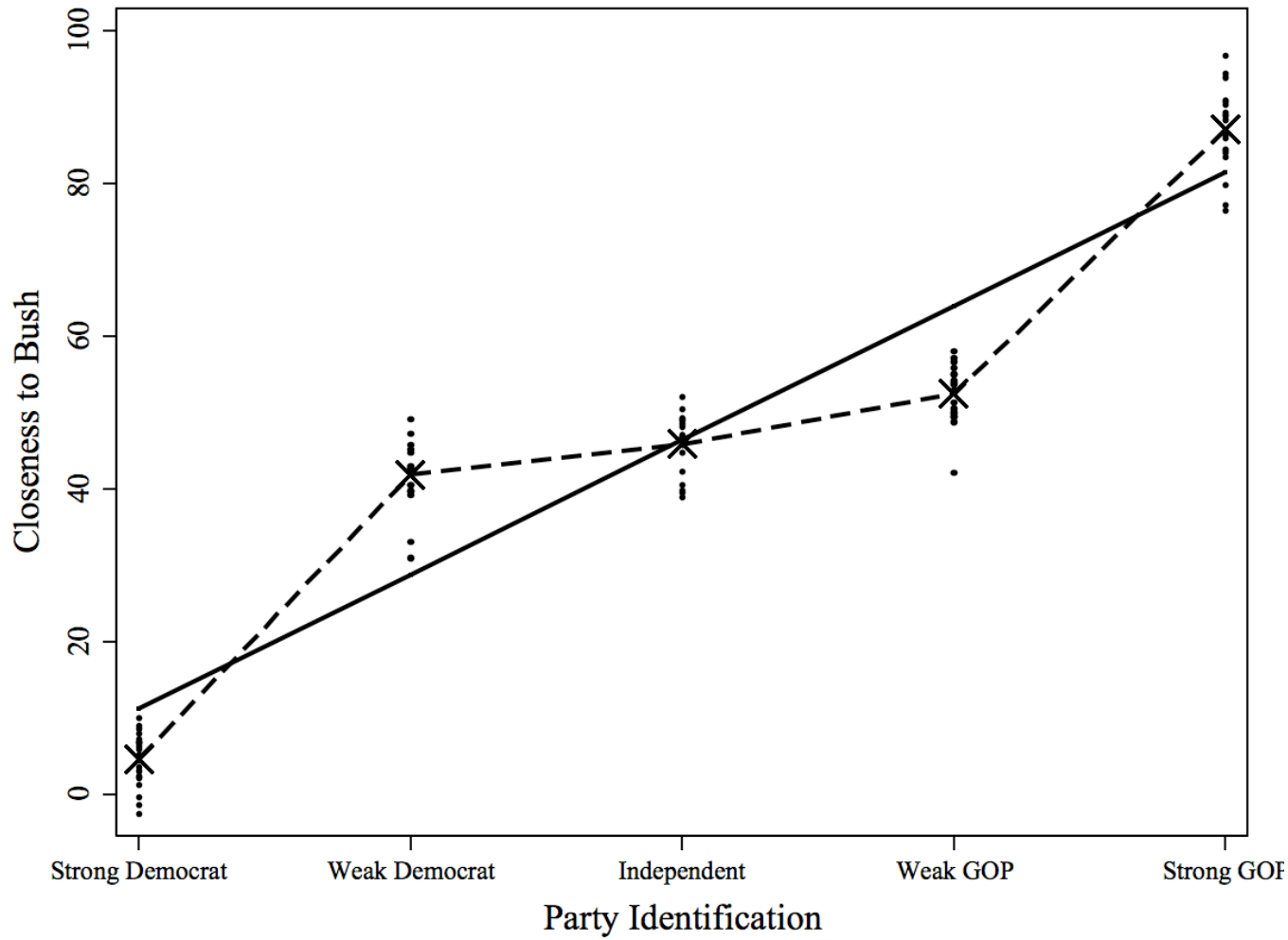
Alternative: "dummy out" `gopscale`:

$$\begin{aligned} \texttt{closeness}_i &= \beta_0 + \beta_1(\texttt{strongdem}_i) + \beta_2(\texttt{weakdem}_i) + \\ &\quad \beta_3(\texttt{weakgop}_i) + \beta_4(\texttt{stronggop}_i) + u_i \end{aligned}$$

yielding:

$$\begin{aligned} \texttt{closeness}_i &= 45.5 - 40(\texttt{strongdem}_i) - 6(\texttt{weakdem}_i) + \\ &\quad 7(\texttt{weakgop}_i) + 42(\texttt{stronggop}_i) + u_i \end{aligned}$$

Ordinal, Illustrated



Dichotomous + Continuous X

E.g.,

$$Y_i = \beta_0 + \beta_1 D_i + \beta_2 X_i + u_i$$

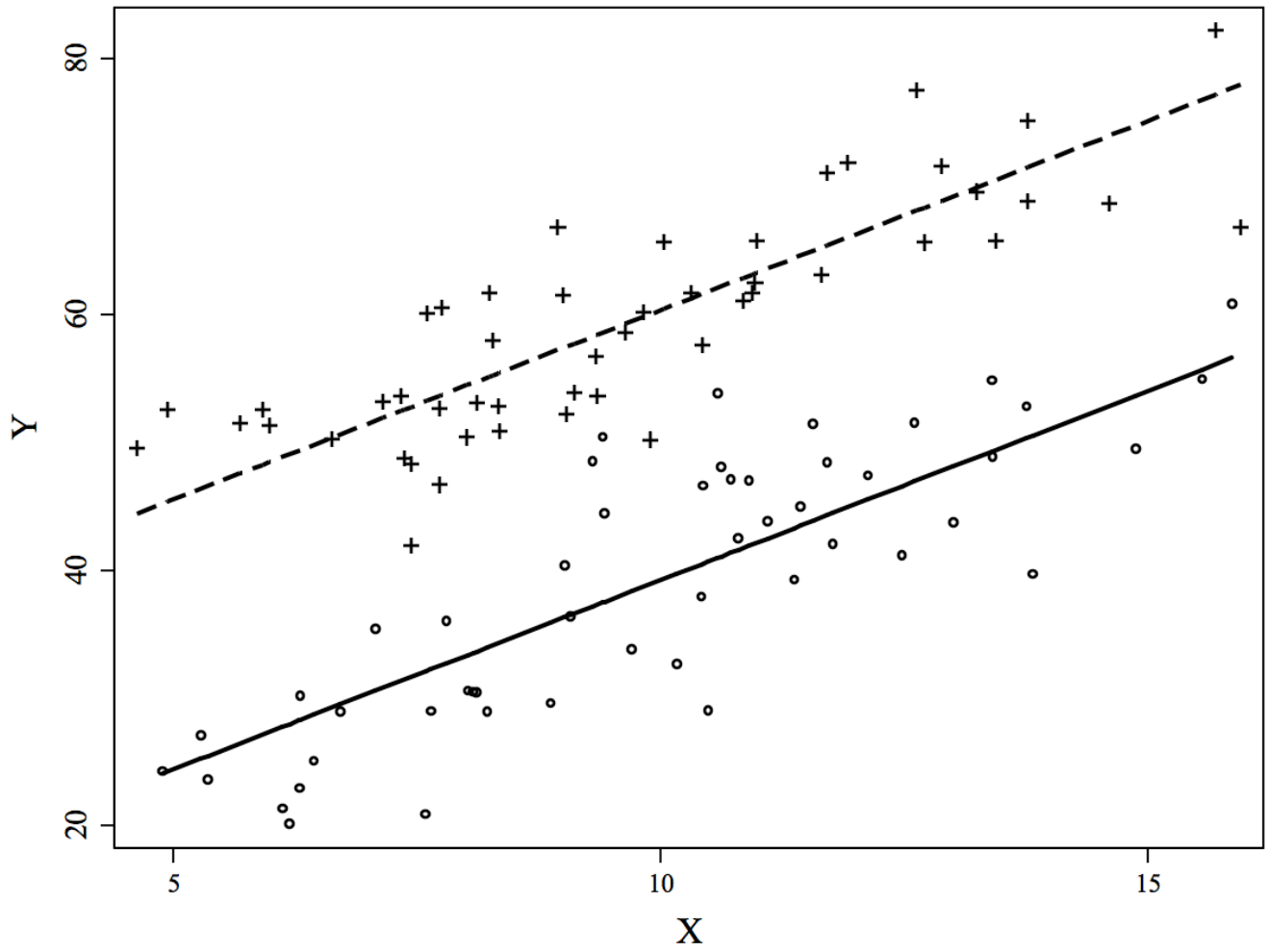
we have

$$\mathrm{E}(Y|X, D = 0) = \beta_0 + \beta_2 X_i$$

and

$$\mathrm{E}(Y|X, D = 1) = (\beta_0 + \beta_1) + \beta_2 X_i.$$

Dichotomous + Continuous X



Examples: SCOTUS (OT1953-1985

```
> summary(SCOTUS)
       id             term           Namici          lctdiss          multlaw
 Min.   :   1   Min.   :53.00   Min.   : 0.000   Min.   :0.0000   Min.   :0.0000
 1st Qu.:1791   1st Qu.:64.00   1st Qu.: 0.000   1st Qu.:0.0000   1st Qu.:0.0000
 Median :3581   Median :72.00   Median : 0.000   Median :0.0000   Median :0.0000
 Mean   :3581   Mean   :71.12   Mean   : 0.842   Mean   :0.1509   Mean   :0.1490
 3rd Qu.:5371   3rd Qu.:79.00   3rd Qu.: 1.000   3rd Qu.:0.0000   3rd Qu.:0.0000
 Max.   :7161   Max.   :85.00   Max.   :39.000   Max.   :1.0000   Max.   :1.0000
                NA's   : 4.00                    NA's   :4.0000   NA's   :5.0000
    civlibs           econs           constit           lctlib
 Min.   :0.0000   Min.   :0.0000   Min.   :0.0000   Min.   :  0.0000
 1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:  0.0000
 Median :1.0000   Median :0.0000   Median :0.0000   Median :  0.0000
 Mean   :0.5009   Mean   :0.1709   Mean   :0.2536   Mean   :  0.3742
 3rd Qu.:1.0000   3rd Qu.:0.0000   3rd Qu.:1.0000   3rd Qu.:  1.0000
 Max.   :1.0000   Max.   :1.0000   Max.   :1.0000   Max.   :  1.0000
                                                    NA's   :120.0000
```

Creating Dummies

All civil rights & economics cases:

```
> SCOTUS$civil.econ<-SCOTUS$civlibs + SCOTUS$econs
```

Factors:

```
> SCOTUS$termdummies<-factor(SCOTUS$term)
> is.factor(SCOTUS$termdummies)
[1] TRUE
> summary(SCOTUS$termdummies)
  53   54   55   56   57   58   59   60   61   62   63   64   65   66   67
 126  109  128  162  196  165  157  160  148  189  223  156  187  201  285

  68   69   70   71   72   73   74   75   76   77   78   79   80   81
 207  185  227  262  269  267  223  253  254  244  244  221  255  269

  82   83   84   85 NA's
 277  298  301  309    4
```

Regressions (vs. *t*-tests...)

```
> fit1<-with(SCOTUS, lm(Namici~civlibs))
> summary(fit1)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.91774    0.03661  25.069  < 2e-16 ***
civlibs     -0.15136    0.05173  -2.926  0.00344 **
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 2.189 on 7159 degrees of freedom
Multiple R-squared:  0.001195,Adjusted R-squared:  0.001055
F-statistic: 8.563 on 1 and 7159 DF,  p-value: 0.003442



> with(SCOTUS, t.test(Namici~civlibs))

Welch Two Sample t-test

data:  Namici by civlibs
t = 2.9258, df = 7114.116, p-value = 0.003446
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 0.04995001 0.25277126
sample estimates:
mean in group 0 mean in group 1
      0.9177392       0.7663786
```

Effect Coding

```
> SCOTUS$civlibeffect<-SCOTUS$civlibs
> SCOTUS$civlibeffect[SCOTUS$civlibs==0]<-(-1)
> fit2<-with(SCOTUS, lm(Namici~SCOTUS$civlibeffect))
> summary(fit2)

Call:
lm(formula = Namici ~ SCOTUS$civlibeffect)

Residuals:
   Min     1Q Median     3Q    Max
-0.918 -0.918 -0.766  0.082 38.234

Coefficients:
                    Estimate Std. Error t value Pr(>|t|)
(Intercept)          0.84206    0.02586  32.559  < 2e-16 ***
SCOTUS$civlibeffect -0.07568    0.02586  -2.926  0.00344 **
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 2.189 on 7159 degrees of freedom
Multiple R-squared:  0.001195,Adjusted R-squared:  0.001055
F-statistic: 8.563 on 1 and 7159 DF,  p-value: 0.003442
```

Many D_is

```
> fit3<-with(SCOTUS, lm(Namici~lctdiss+multlaw+civlibs+
+                         econs+constit+lctlib))
> summary(fit3)

Call:
lm(formula = Namici ~ lctdiss + multlaw + civlibs + econs + constit +
    lctlib)

Residuals:
   Min     1Q Median     3Q    Max
-2.582 -0.976 -0.472 -0.260 37.086

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.47245    0.05273   8.960  < 2e-16 ***
lctdiss      0.36760    0.07173   5.125 3.06e-07 ***
multlaw      0.61306    0.07445   8.235  < 2e-16 ***
civlibs     -0.21255    0.06022  -3.530 0.000419 ***
econs        0.08772    0.07652   1.146 0.251691
constit      0.53793    0.06372   8.442  < 2e-16 ***
lctlib       0.50309    0.05396   9.323  < 2e-16 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 2.15 on 7033 degrees of freedom
  (121 observations deleted due to missingness)
Multiple R-squared:  0.05013,Adjusted R-squared:  0.04932
F-statistic: 61.86 on 6 and 7033 DF,  p-value: < 2.2e-16
```

Using factor

```
> fit4<-with(SCOTUS, lm(Namici~lctdiss+multlaw+civlibs+
+                         econs+constit+lctlib+term))
> summary(fit4)

Call:
lm(formula = Namici ~ lctdiss + multlaw + civlibs + econs + constit +
    lctlib + term)

Residuals:
   Min     1Q Median     3Q    Max
-2.968 -0.906 -0.428  0.143 36.958

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -2.726962   0.202367 -13.475  < 2e-16 ***
lctdiss      0.359494   0.070415   5.105 3.39e-07 ***
multlaw      0.649932   0.073109   8.890  < 2e-16 ***
civlibs     -0.289314   0.059295  -4.879 1.09e-06 ***
econs        0.199464   0.075419   2.645  0.00819 **
constit      0.515435   0.062559   8.239  < 2e-16 ***
lctlib       0.339891   0.053901   6.306 3.04e-10 ***
term         0.046142   0.002821  16.354  < 2e-16 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 2.11 on 7032 degrees of freedom
  (121 observations deleted due to missingness)
Multiple R-squared:  0.08493,Adjusted R-squared:  0.08402
F-statistic: 93.24 on 7 and 7032 DF,  p-value: < 2.2e-16
```

Using factor

```
> fit5<-with(SCOTUS, lm(Namici~lctdiss+multlaw+civlibs+
+                       econs+constit+lctlib+as.factor(term)))
> summary(fit5)

Call:
lm(formula = Namici ~ lctdiss + multlaw + civlibs + econs + constit +
    lctlib + as.factor(term))

Residuals:
   Min     1Q Median     3Q    Max
-3.064 -0.920 -0.384  0.106 36.831

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)       -0.16153    0.19530  -0.827 0.408200
lctdiss            0.34558    0.07067   4.890 1.03e-06 ***
multlaw            0.64348    0.07334   8.774  < 2e-16 ***
civlibs           -0.27137    0.05967  -4.548 5.51e-06 ***
econs              0.20039    0.07581   2.643 0.008232 **
constit            0.54280    0.06297   8.620  < 2e-16 ***
lctlib             0.33863    0.05458   6.205 5.80e-10 ***
.
.
.
```

Using `factor` (continued)

```
as.factor(term)54  0.26276    0.27934   0.941 0.346918
as.factor(term)55  0.20958    0.26804   0.782 0.434309
as.factor(term)56  0.12536    0.25126   0.499 0.617859
as.factor(term)57  0.06432    0.24227   0.265 0.790654
as.factor(term)58  0.08353    0.25274   0.331 0.741025
.
.
.
as.factor(term)71  0.62313    0.23019   2.707 0.006806 **
as.factor(term)72  0.59503    0.22929   2.595 0.009476 **
as.factor(term)73  0.78179    0.22918   3.411 0.000650 ***
as.factor(term)74  0.53254    0.23636   2.253 0.024287 *
as.factor(term)75  0.80353    0.23118   3.476 0.000513 ***
as.factor(term)76  0.49269    0.23138   2.129 0.033262 *
as.factor(term)77  1.07725    0.23265   4.630 3.72e-06 ***
as.factor(term)78  1.04335    0.23243   4.489 7.27e-06 ***
as.factor(term)79  0.85363    0.23696   3.602 0.000318 ***
as.factor(term)80  1.21205    0.23183   5.228 1.76e-07 ***
as.factor(term)81  1.49347    0.22925   6.515 7.80e-11 ***
as.factor(term)82  1.46004    0.22858   6.388 1.79e-10 ***
as.factor(term)83  1.29417    0.22549   5.739 9.90e-09 ***
as.factor(term)84  1.23434    0.22517   5.482 4.36e-08 ***
as.factor(term)85  1.59037    0.22491   7.071 1.68e-12 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 2.108 on 7001 degrees of freedom
  (121 observations deleted due to missingness)
Multiple R-squared:  0.0914,Adjusted R-squared:  0.08647
F-statistic: 18.53 on 38 and 7001 DF,  p-value: < 2.2e-16
```

`factor` results, plotted (1953 = 0)