Regression with interaction terms

PSCI 2301: Quantitative Political Science II

Prof. Brenton Kenkel

brenton.kenkel@gmail.com

Vanderbilt University

February 26, 2025

Recap

Last time — regression for causal inference

1. The linear model with controls
    - Reduce standard errors by isolating treatment effect
    - Control for selection bias
2. Interpreting regression output
    - “All else equal” comparison
    - Don’t try to interpret control variable coefficients

Today's agenda

Regression with interaction terms

1. Heterogeneous effects and why we might look for them
2. Interaction with a categorical variable
3. Interaction with a continuous variable
    - Linear interaction approach
    - Discretization approach

Heterogeneous effects and the interactive model

Heterogeneous effects: What they are, why we care

We've mostly focused on overall average treatment effects

Sometimes we suspect different average effects across subpopulations

→ e.g., does social pressure affect turnout differently for men and women?

Reasons to look for these **heterogeneous effects**

- Theoretical
  - → Think about mechanisms — why you expect treatment to be effective
  - → Would the mechanism be stronger/weaker for some groups?
- Practical
  - → How can the treatment be targeted most effectively?
  - → e.g., can we get most of the votes for a fraction of the cost?

Heterogeneous effects in randomized experiments

Potential outcomes Y_{1i}, Y_{0i}

Treatment indicator D_i

Binary group indicator X_i

→ e.g., $X_i = 1$ for men, $X_i = 0$ for women

Overall average treatment effect $\mathbb{E}[Y_{1i} - Y_{0i}]$

Average treatment effect among men $\mathbb{E}[Y_{1i} - Y_{0i} \mid X_i = 1]$

Average treatment effect among women $\mathbb{E}[Y_{1i} - Y_{0i} \mid X_i = 0]$

→ same subgroup effects we calculate for subclassification, now just treating them as interesting on their own

Capturing heterogeneous effects with a linear model

Regression with an interaction term:

$$\mathbb{E}[Y_i \mid D_i, Z_i] = \alpha + \beta_1 D_i + \beta_2 X_i + \underbrace{\beta_3 (D_i \times X_i)}_{\text{interaction term}}$$

- α: average outcome in control condition among $X_i = 0$ subgroup
  - → voter turnout among women who don't receive the GOTV mailer
- β_1: average treatment effect in $X_i = 0$ subgroup
  - → average effect of mailer on turnout among women
- β_2: difference in avg outcome in control cond. b/w $X_i = 1$ and $X_i = 0$
  - → among those w/ no mailer, how much more/less is men's turnout?
- β_3 difference in ATE b/w $X_i = 1$ and $X_i = 0$
  - → how much more/less does mailer affect turnout among men?

Recap: The GGL turnout data

```r
print(df_ggl)
```

```
# A tibble: 344,084 × 16
  sex      yob g2000 g2002 g2004 p2000 p2002 p2004 treatment cluster voted hh_id
  <chr>  <dbl> <chr> <chr> <chr> <chr> <chr> <chr> <chr>       <dbl> <dbl> <dbl>
1 male    1941 yes   yes   yes   no    yes   No    Civic Du…       1     0     1
2 female  1947 yes   yes   yes   no    yes   No    Civic Du…       1     0     1
3 male    1951 yes   yes   yes   no    yes   No    Hawthorne       1     1     2
4 female  1950 yes   yes   yes   no    yes   No    Hawthorne       1     1     2
5 female  1982 yes   yes   yes   no    yes   No    Hawthorne       1     1     2
# ℹ 344,079 more rows
# ℹ 4 more variables: hh_size <dbl>, numberofnames <dbl>, p2004_mean <dbl>,
#   g2004_mean <dbl>
```

To begin — just look at Neighbors treatment vs no mailer

```r
df_ggl_subset <- df_ggl |>
  filter(treatment %in% c("Neighbors", "Control")) |>
  mutate(treat = if_else(treatment == "Neighbors", 1, 0))
```

Estimating an interactive model in R

```
fit_interact <- lm(voted ~ treat * sex, data = df_ggl_subset)
tidy(fit_interact)
```

```
# A tibble: 4 × 5
  term          estimate std.error statistic   p.value
  <chr>            <dbl>     <dbl>     <dbl>     <dbl>
1 (Intercept)   0.290      0.00149   194.     0
2 treat         0.0809     0.00366    22.1    3.02e-108
3 sexmale       0.0123     0.00211     5.85   5.05e-  9
4 treat:sexmale 0.000849   0.00517     0.164  8.70e-  1
```

Term	Value	Interpretation
(Intercept)	0.290	29% turnout among women in Control group
treat	0.081	8.1% ATE among women
sexmale	0.012	1.2% higher turnout among men in Control group (i.e., 30.2%)
treat:sexmale	0.0008	0.08% higher ATE among men (i.e., 8.18%)

Another discrete interaction example

Divide into three categories based on household size:

```r
df_ggl_subset <- df_ggl_subset |>
  mutate(hh_type = case_when(
          hh_size == 1 ~ "Single",
          hh_size == 2 ~ "Couple",
          TRUE ~ "More than two"))

fit_hh_type <- lm(voted ~ treat * hh_type, data = df_ggl_subset)
tidy(fit_hh_type)
```

```
# A tibble: 6 × 5
  term                       estimate std.error statistic   p.value
  <chr>                         <dbl>     <dbl>     <dbl>     <dbl>
1 (Intercept)                 0.303     0.00134   227.    0        
2 treat                       0.0878    0.00328    26.8   4.29e-158
3 hh_typeMore than two       -0.0423    0.00254   -16.7   1.64e- 62
4 hh_typeSingle               0.0277    0.00313     8.84  9.54e- 19
5 treat:hh_typeMore than two -0.0303    0.00622    -4.87  1.15e-  6
6 treat:hh_typeSingle         0.00418   0.00764     0.547 5.84e-  1
```

Heterogeneous effects across a continuous variable

What if we wanted to see whether the treatment effect varies by age?

Linear interactive effect model

$$\mathbb{E}[Y_i \mid D_i, X_i] = \alpha + \beta_1 D_i + \beta_2 X_i + \beta_3 (D_i \times X_i)$$

Same equation as before, but new interpretation

Average treatment effect when $X_i = x$:

$$\mathbb{E}[Y_i \mid D_i = 1, X_i = x] - \mathbb{E}[Y_i \mid D_i = 0, X_i = x] = \beta_1 + \beta_3 x$$

β_3 = how much does treatment effect increase/decrease with each unit of X_i?

Heterogeneous effects across a continuous variable

```
fit_interact_age <- lm(voted ~ treat * yob, data = df_ggl_subset)
tidy(fit_interact_age)
```

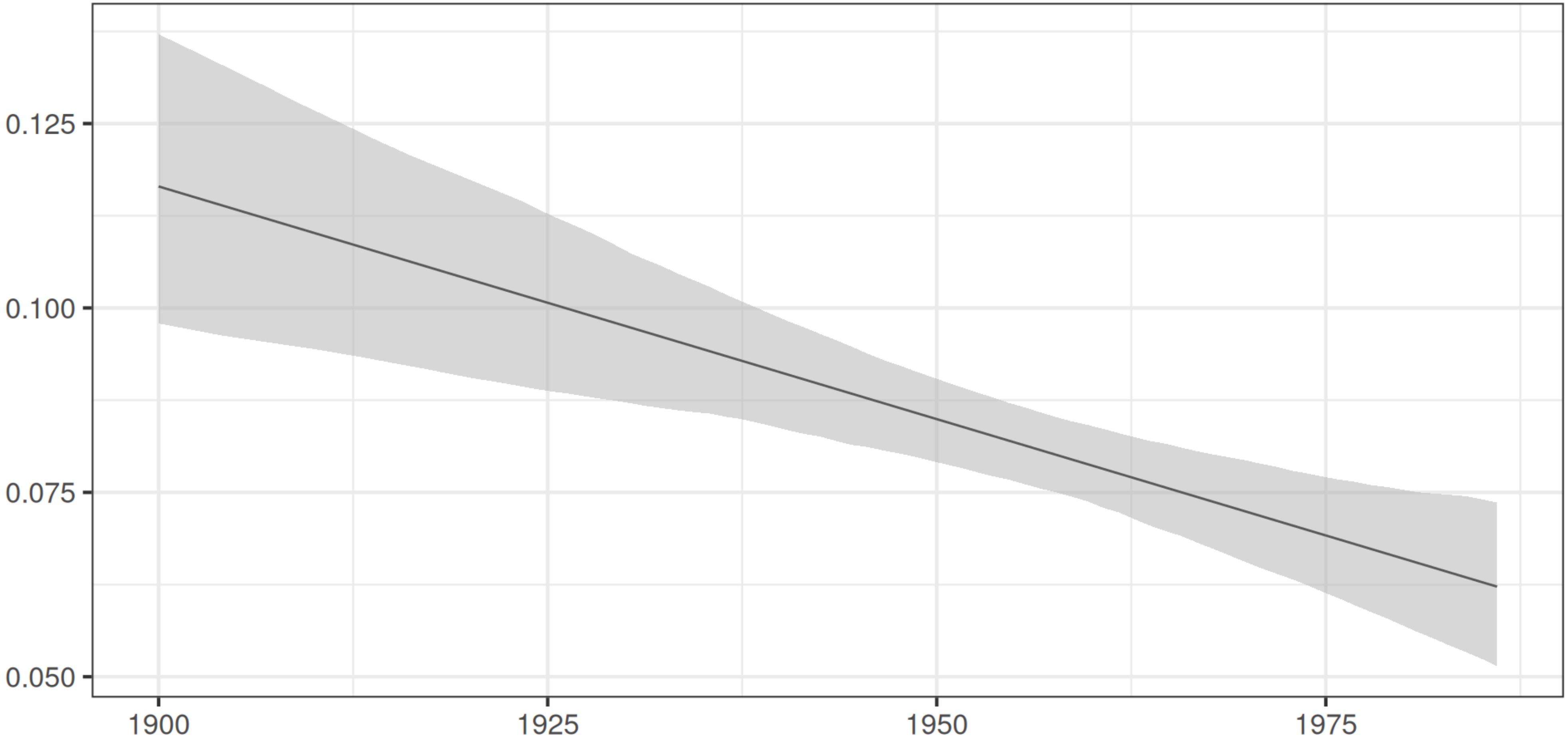
```
# A tibble: 4 × 5
  term         estimate std.error statistic  p.value
  <chr>           <dbl>     <dbl>     <dbl>    <dbl>
1 (Intercept)  8.12     0.142          57.2  0
2 treat        1.31     0.345           3.80 0.000145
3 yob         -0.00400  0.0000725     -55.1  0
4 treat:yob   -0.000628 0.000176       -3.57 0.000364
```

ATE = 1.31 - 0.000628 $\times$ birth year

- born 1925: 10.11% turnout increase
- born 1950: 8.54% turnout increase
- born 1975: 6.97% turnout increase

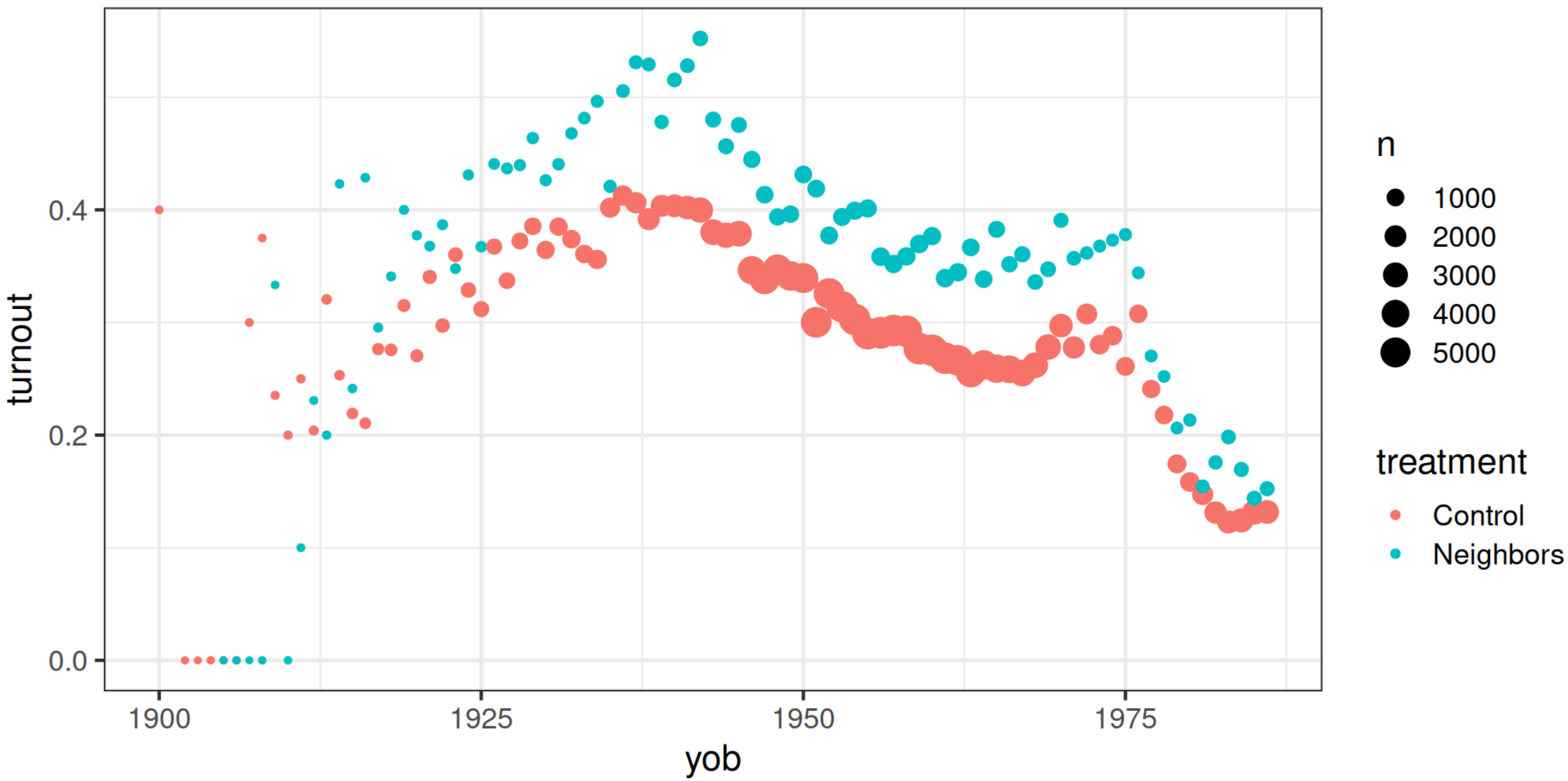
Heterogeneous effects across a continuous variable

```r
library("interplot")
interplot(m = fit_interact_age, var1 = "treat", var2 = "yob")
```



Heterogeneous effects across a continuous variable

Problem with the standard approach — is ATE linear in age?



Heterogenous effects across a continuous variable

Another approach — discretize the continuous variable

```r
df_ggl_subset <- df_ggl_subset |>
  mutate(yob_cat = cut(yob - 1900, breaks = seq(0, 90, by = 10)))

df_ggl_subset |> relocate(yob_cat, yob)
```

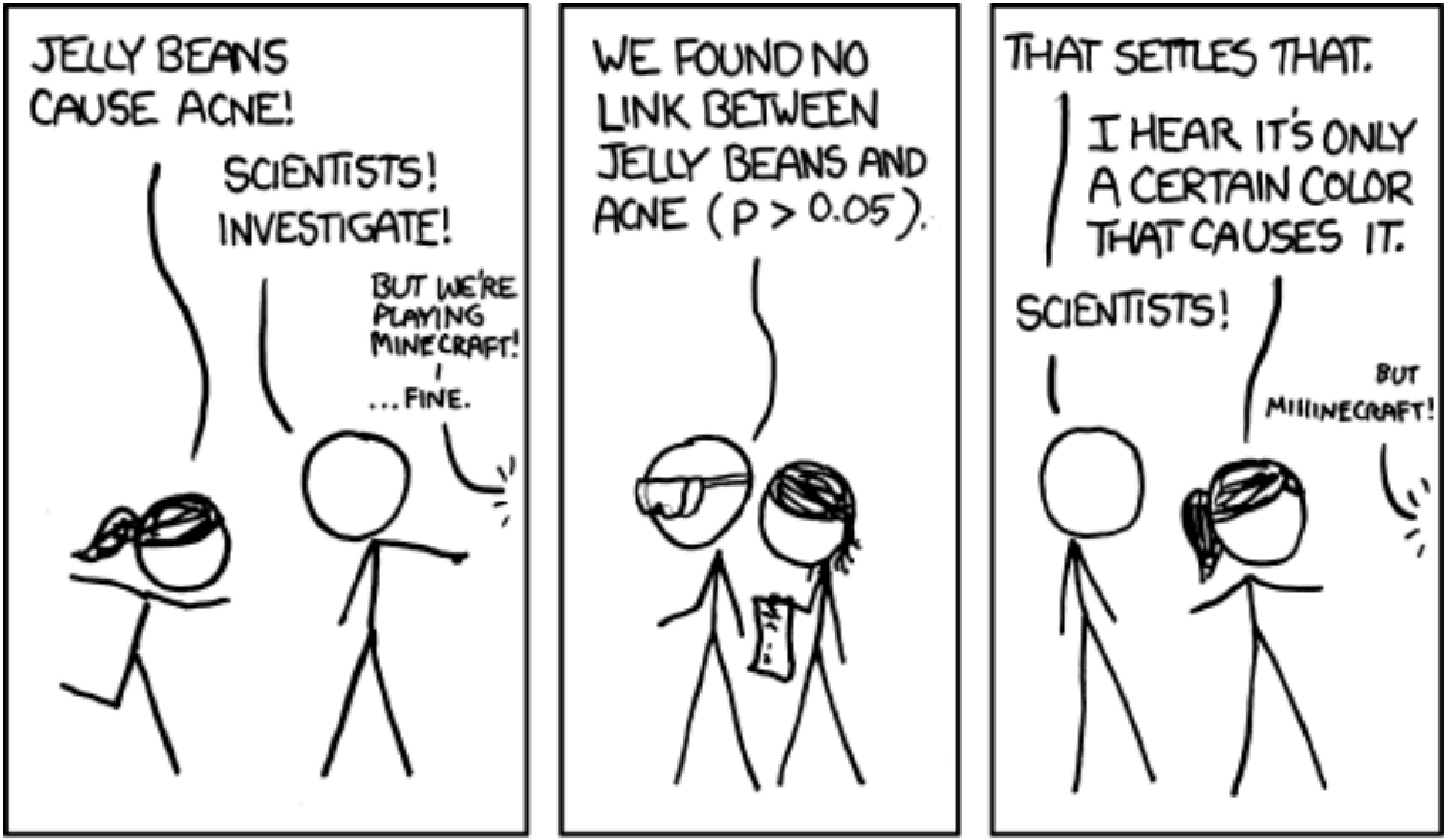
```
# A tibble: 229,444 × 19
  yob_cat   yob sex    g2000 g2002 g2004 p2000 p2002 p2004 treatment cluster
  <fct>   <dbl> <chr>  <chr> <chr> <chr> <chr> <chr> <chr> <chr>       <dbl>
1 (80,90]  1981 male   no    no    yes   no    no    No    Control         1
2 (50,60]  1959 female yes   yes   yes   no    yes   No    Control         1
3 (50,60]  1956 male   yes   yes   yes   no    yes   No    Control         1
4 (60,70]  1968 female no    no    yes   no    yes   No    Control         1
5 (60,70]  1967 male   yes   yes   yes   no    yes   No    Control         1
# i 229,439 more rows
# i 8 more variables: voted <dbl>, hh_id <dbl>, hh_size <dbl>,
#   numberofnames <dbl>, p2004_mean <dbl>, g2004_mean <dbl>, treat <dbl>,
#   hh_type <chr>
```

Heterogeneous effects across a continuous variable

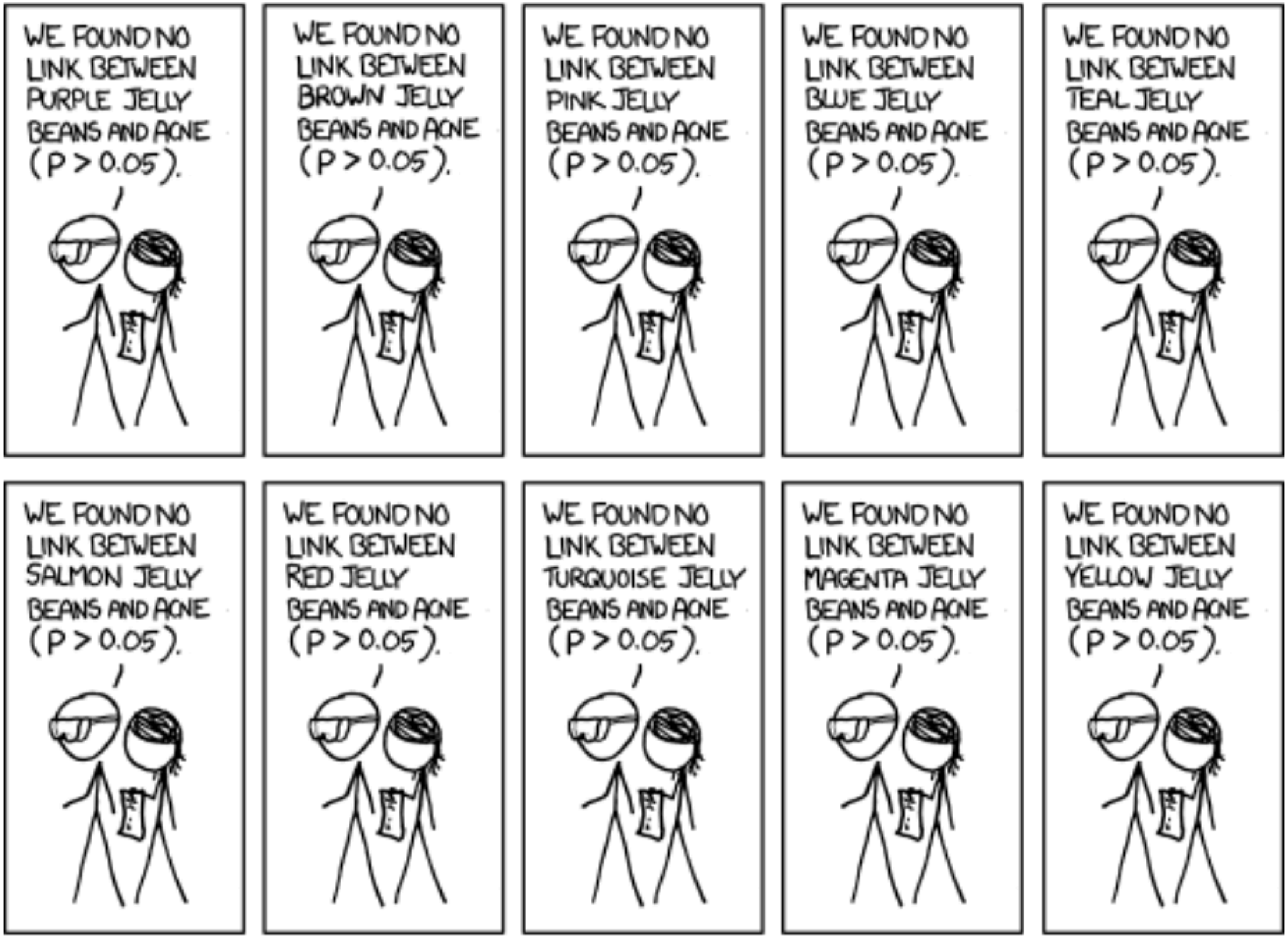
```r
fit_yob_cat <- lm(voted ~ treat * yob_cat, data = df_ggl_subset)
tidy(fit_yob_cat)
```

```
# A tibble: 18 × 5
   term                 estimate std.error statistic  p.value
   <chr>                   <dbl>     <dbl>     <dbl>    <dbl>
 1 (Intercept)            0.197     0.0542     3.64  0.000275
 2 treat                 -0.0795    0.123     -0.645 0.519
 3 yob_cat(10,20]         0.0712    0.0554     1.28  0.199
 4 yob_cat(20,30]         0.154     0.0545     2.83  0.00464
 5 yob_cat(30,40]         0.196     0.0543     3.60  0.000319
 6 yob_cat(40,50]         0.163     0.0543     3.00  0.00268
 7 yob_cat(50,60]         0.0986    0.0542     1.82  0.0690
 8 yob_cat(60,70]         0.0677    0.0542     1.25  0.212
 9 yob_cat(70,80]         0.0574    0.0543     1.06  0.291
10 yob_cat(80,90]        -0.0664    0.0543    -1.22  0.221
11 treat:yob_cat(10,20]   0.147     0.126      1.16  0.246
12 treat:yob_cat(20,30]   0.147     0.124      1.19  0.235
13 treat:yob_cat(30,40]   0.179     0.124      1.45  0.148
```

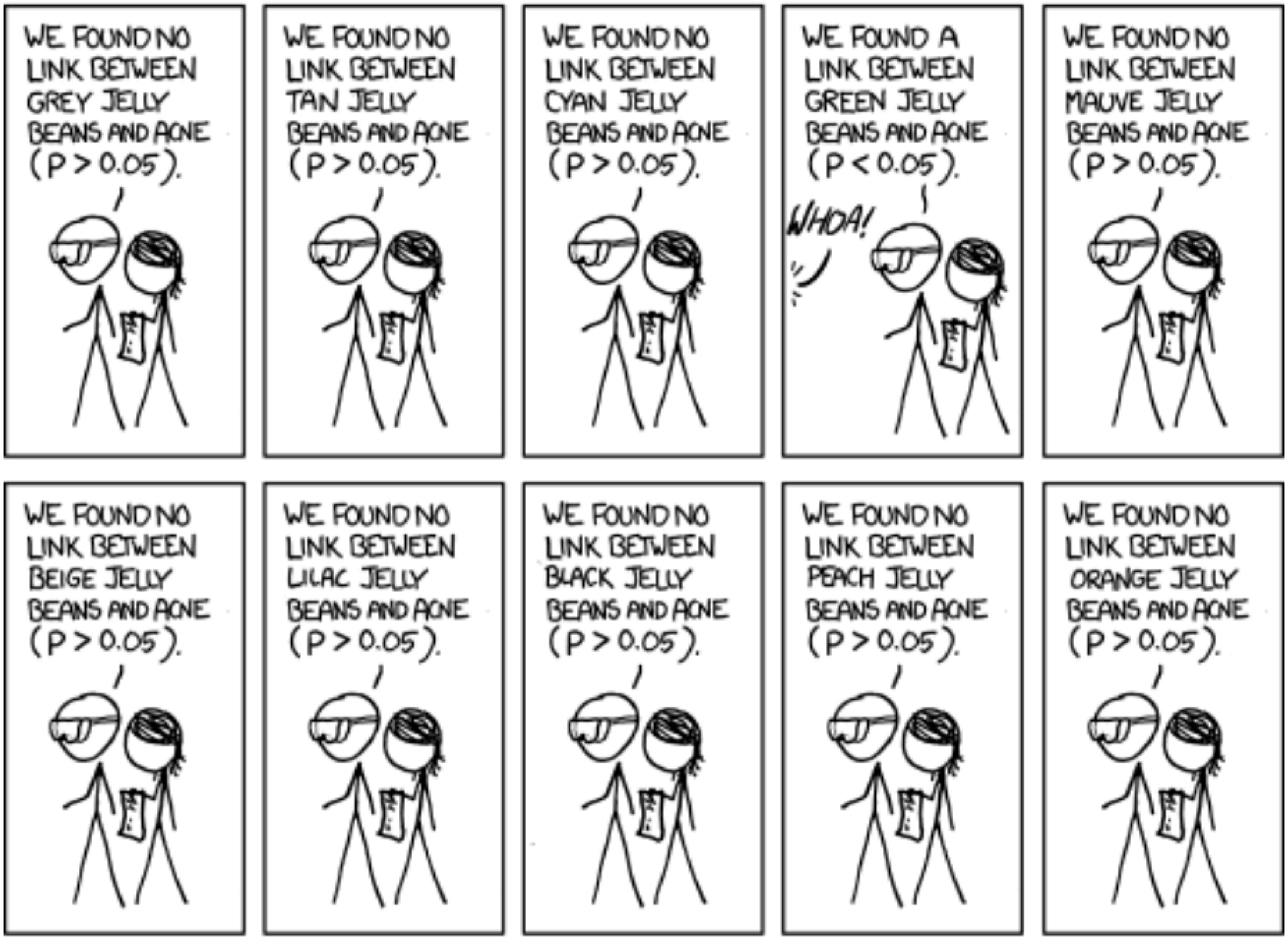
Be careful when looking at lots of interactions!



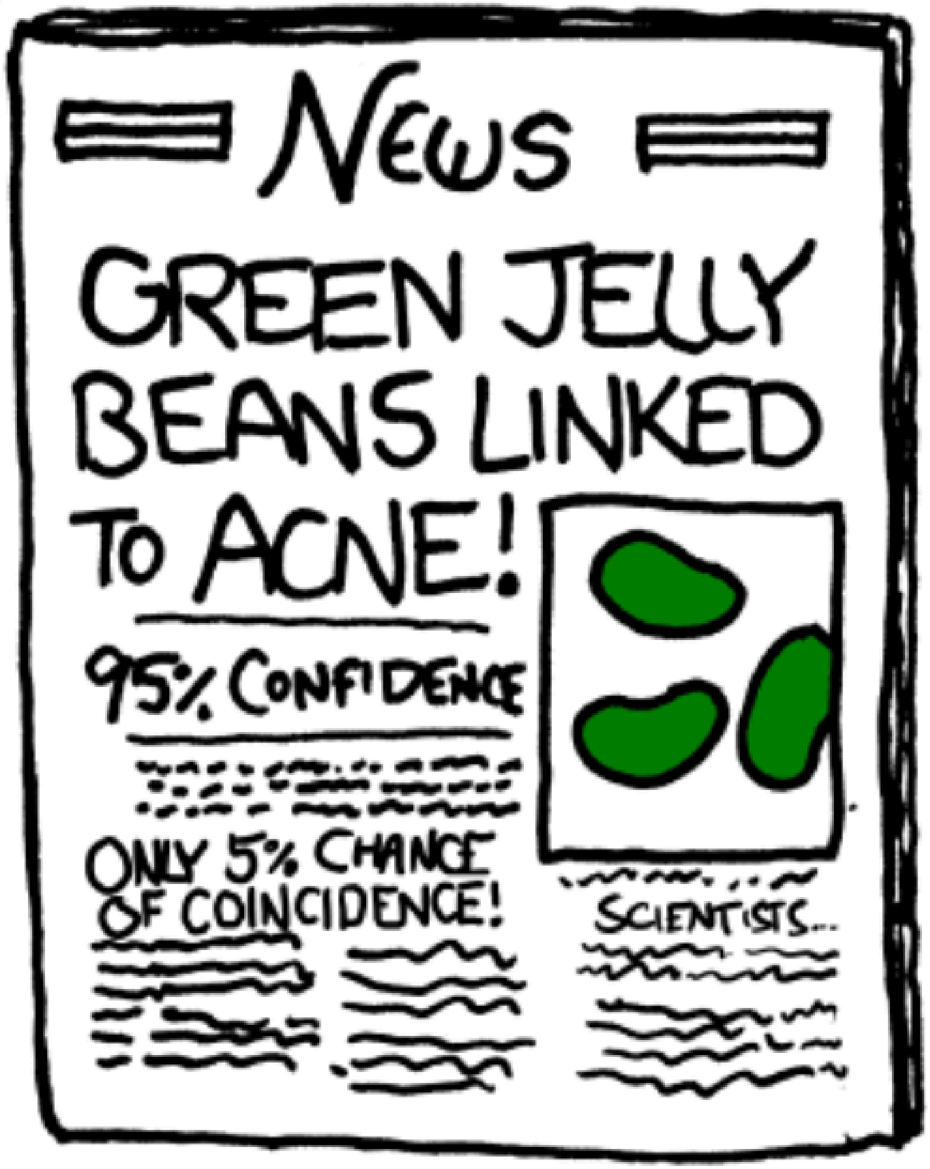
Be careful when looking at lots of interactions!



Be careful when looking at lots of interactions!



Be careful when looking at lots of interactions!



Heterogeneous effects in observational data

What if you're looking for heterogeneous effects when treatment isn't randomly assigned?

Methods are essentially the same — just include controls alongside the interaction

You *can* include multiple interactions but interpretation gets tricky

Wrapping up

What we did today

1. Reasons to look for heterogeneous effects
    - Theoretical: where we expect mechanisms to be stronger/weaker
    - Practical: targeting implementation to get max effect
2. Interpreting interactive regressions
    - Coefficient on interaction term = *difference* in treatment effects
    - Special considerations for interaction w/ continuous variable

Next week

What if we can't measure and control for every plausible confounder?

One way to deal with unobserved confounding — instrumental variables

→ Essentially, random *influences* on non-random treatment assignment

1. Turn in project proposal this Friday
2. Read Acemoglu, Johnson, Robinson, "The Colonial Origins of Economic Development"
3. Read chapter 3 of *Mastering 'Metrics*