Poisson regression coefficients

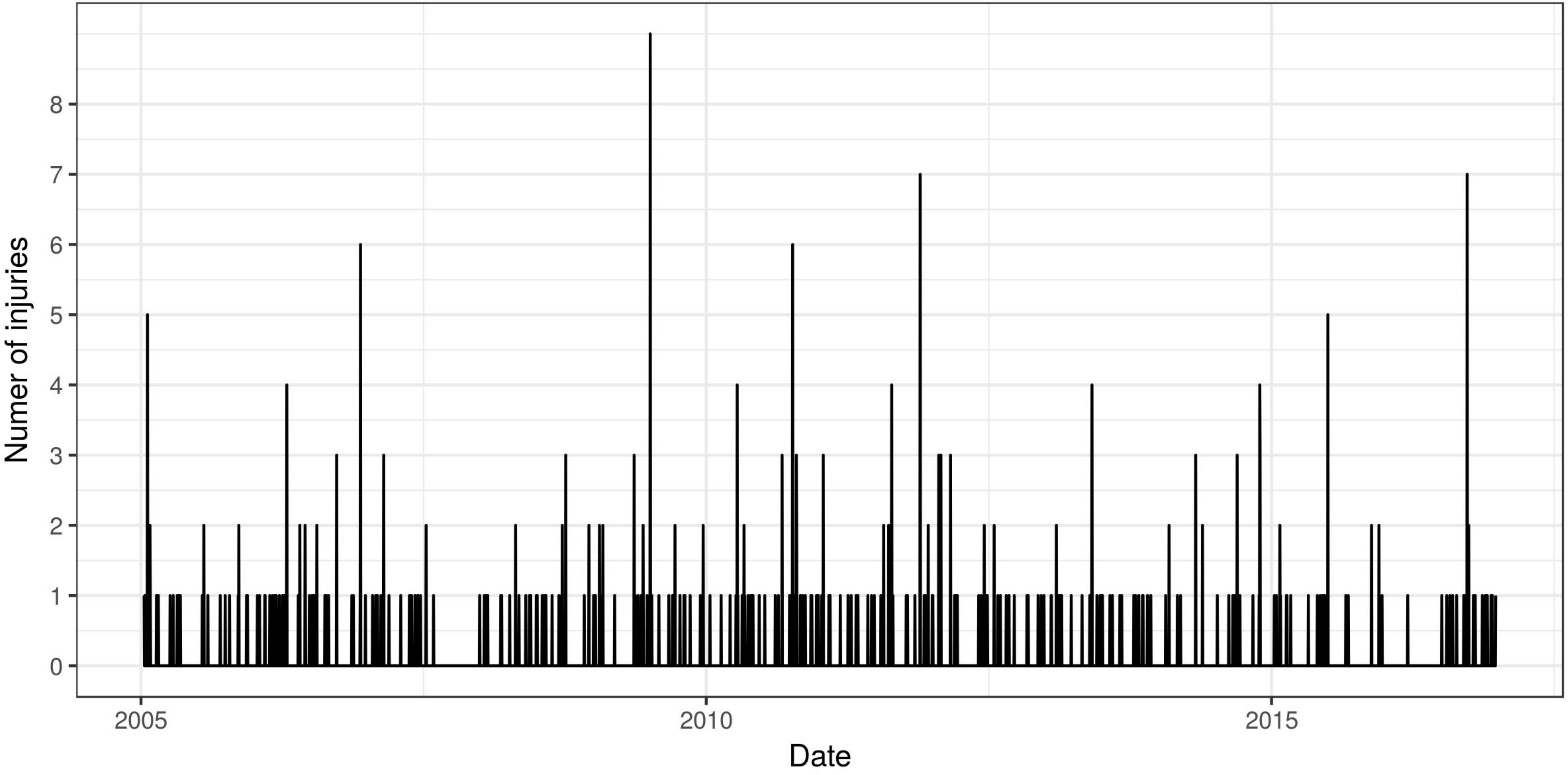
GENERALIZED LINEAR MODELS IN R

Richard Erickson
Instructor

Chapter overview

- Describing Poisson regressions
- Plotting Poisson GLMs with ggplot2
- Describing logistic regression with odds-ratios
- Plotting binomial GLMs with ggplot2

Fire injury data



Linear model coefficients overview

- Estimate expected daily injury per month
- Estimate reference intercept
- Estimate intercept for other months

Linear model equation

- $y \sim \beta_0 + \beta_m x_m + \ldots + \epsilon$
- β_0: Reference intercept
- β_m: Month m effect
- y injuries per day (e.g., 1, 0, 4)
- x dummy variable to code for month (0 or 1)
- m corresponds to month intercept, dummy variable

Linear model results

- β_0 is expected (or average) in reference month
- β_m is effect of month m (or difference from reference)
- e.g., β_0 + β_m = ave. daily injuries for month m
- More complicated models covered in chapter 4
- **Linear models are additive**

Poisson model

- $y \sim \text{Poisson}(\lambda)$
- **Link:** $\lambda = e^{(\beta_0 + \beta_m x_m + \epsilon)}$
- **Multiplicative**
- Example results:
  - $\beta_0 \times \beta_1 = \ln(\text{mean daily injuries for month } m)$
  - Take exponential to convert to raw units

Difference between Poisson and linear models

- **Poisson model:** $e^{\beta_0 \times \beta_m}$ = expected daily injuries for month m
- **Linear model:** β_0 + β_m = expected daily injuries for month m

Extract in R

```
poissonOut <- glm(y ~ x, family = 'poisson')
coef(poissonOut)
exp(coef(poissonOut))
```

Tidy solution

```
library(broom)
poissonOut <- glm(y ~ x, family = 'poisson')
tidy(poissonOut, exponentiate = TRUE)
```

Statistical inferences

- Similar as linear model on link-scale
- Do coefficients differ from zero?
- On data-scale, different
- Do coefficients differ from 1?
- **Exponential**-scale rather than **raw**-scale

Let's practice!

GENERALIZED LINEAR MODELS IN R

Plotting Poisson regression

GENERALIZED LINEAR MODELS IN R

Richard Erickson
Instructor

When to use geom_smooth with Poisson

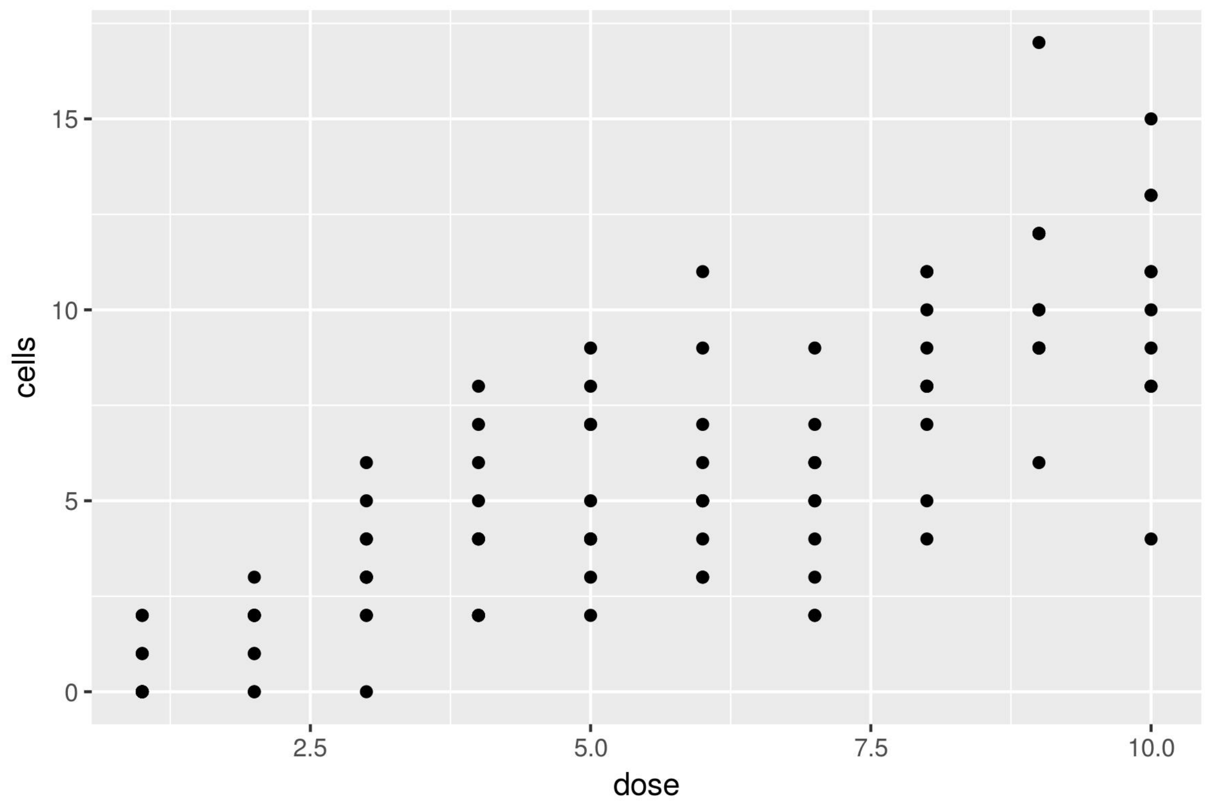
- Works best with continuous predictor variables
- e.g., increasing dose and number of cells with cancer per cm^2
- Otherwise, use boxplot or similar plotting tool

Cancer cells dose study

- Simulate data
- Dose-response
  - x: Dose
  - y: Number of cancer cells per cm^2

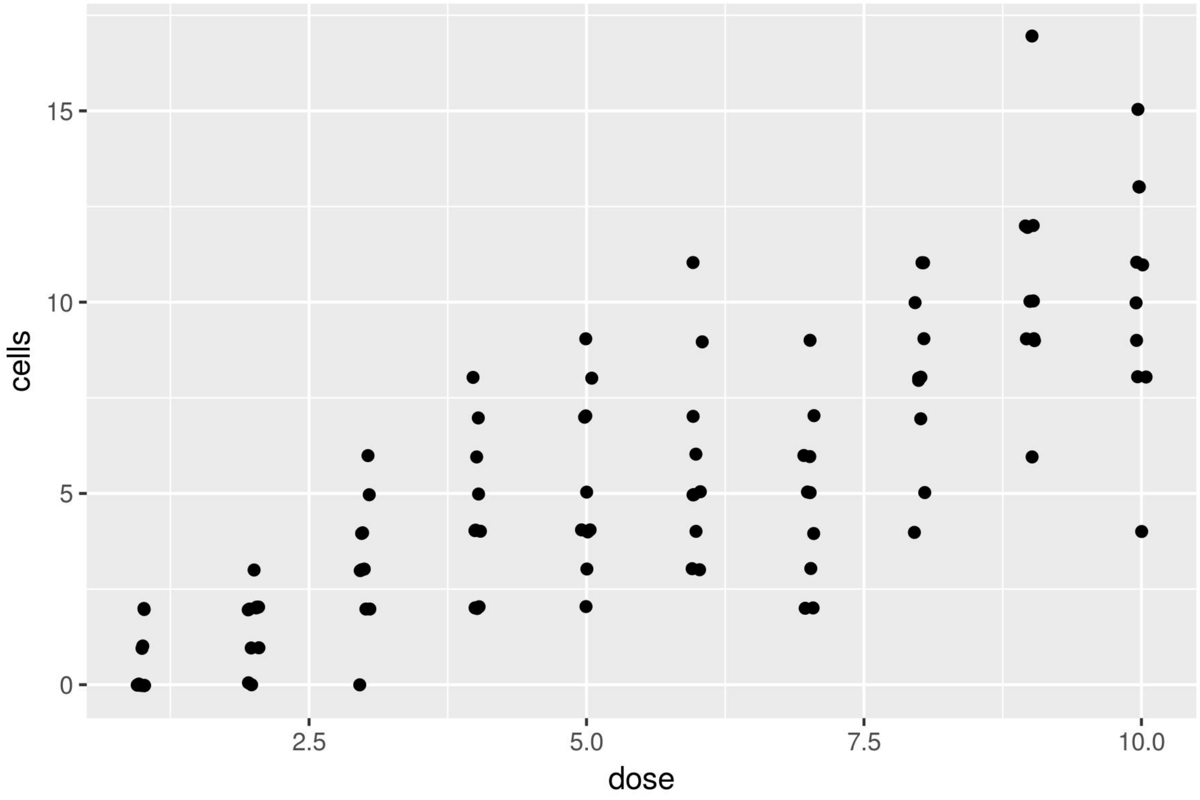
Plot points

```r
ggplot(data = dat, aes(x = dose, y = cells)) +
    geom_point()
```



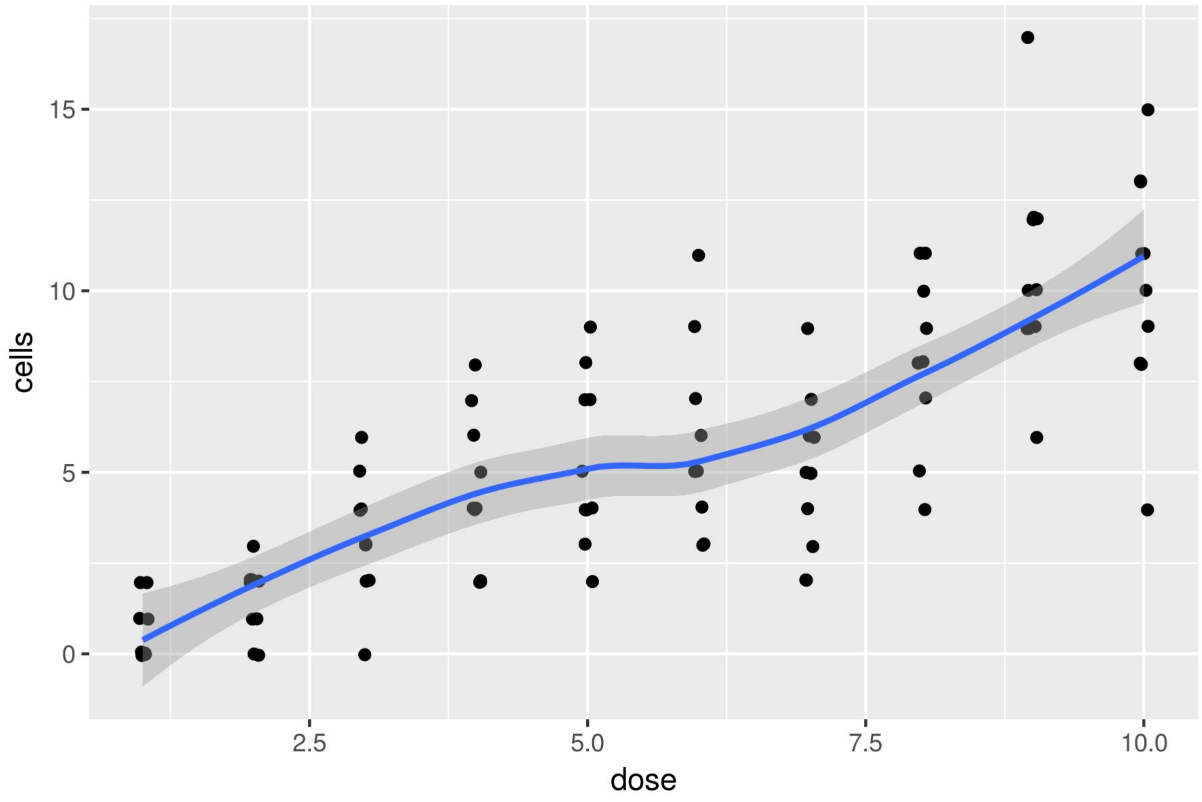
Jitter points

```
ggplot(data = dat, aes(x = dose, y = cells)) +
   geom_jitter(width = 0.05, height = 0.05)
```



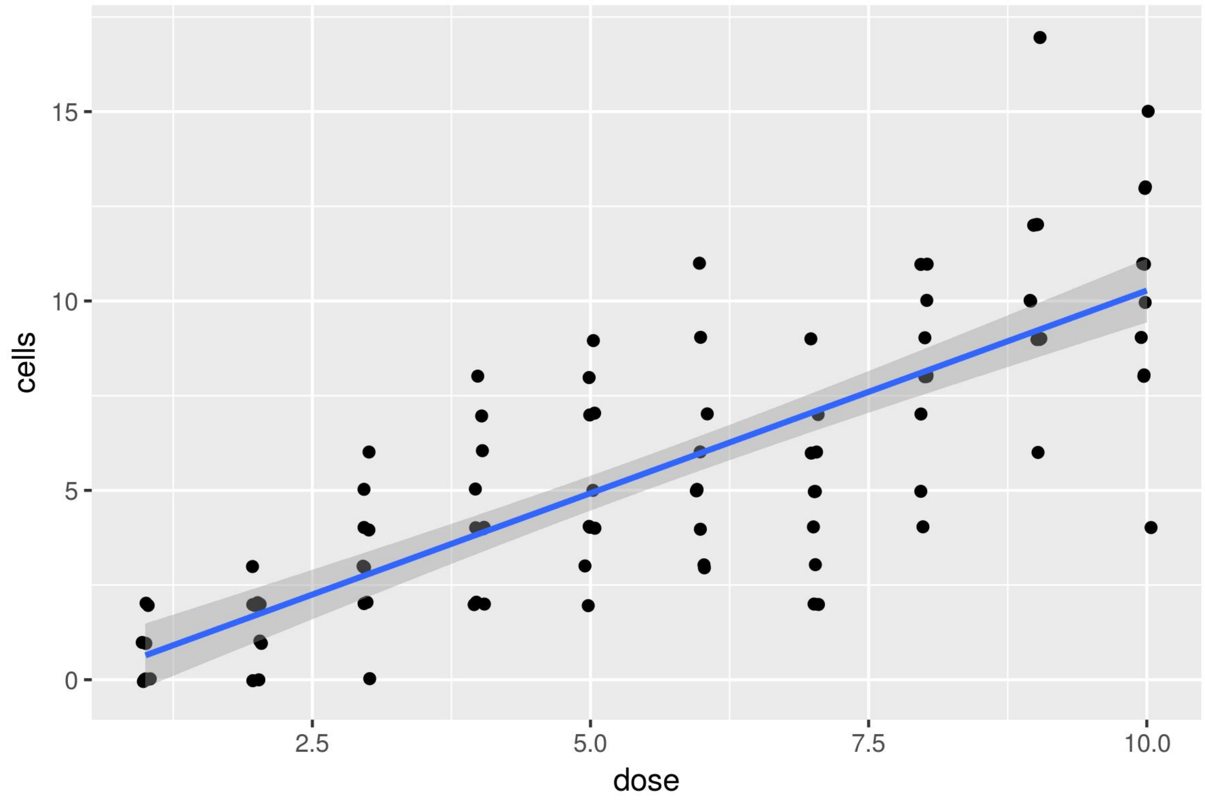
geom_smooth()

```
ggplot(data = dat, aes(x = dose, y = cells)) +
   geom_jitter(width = 0.05, height = 0.05)
   geom_smooth()
```



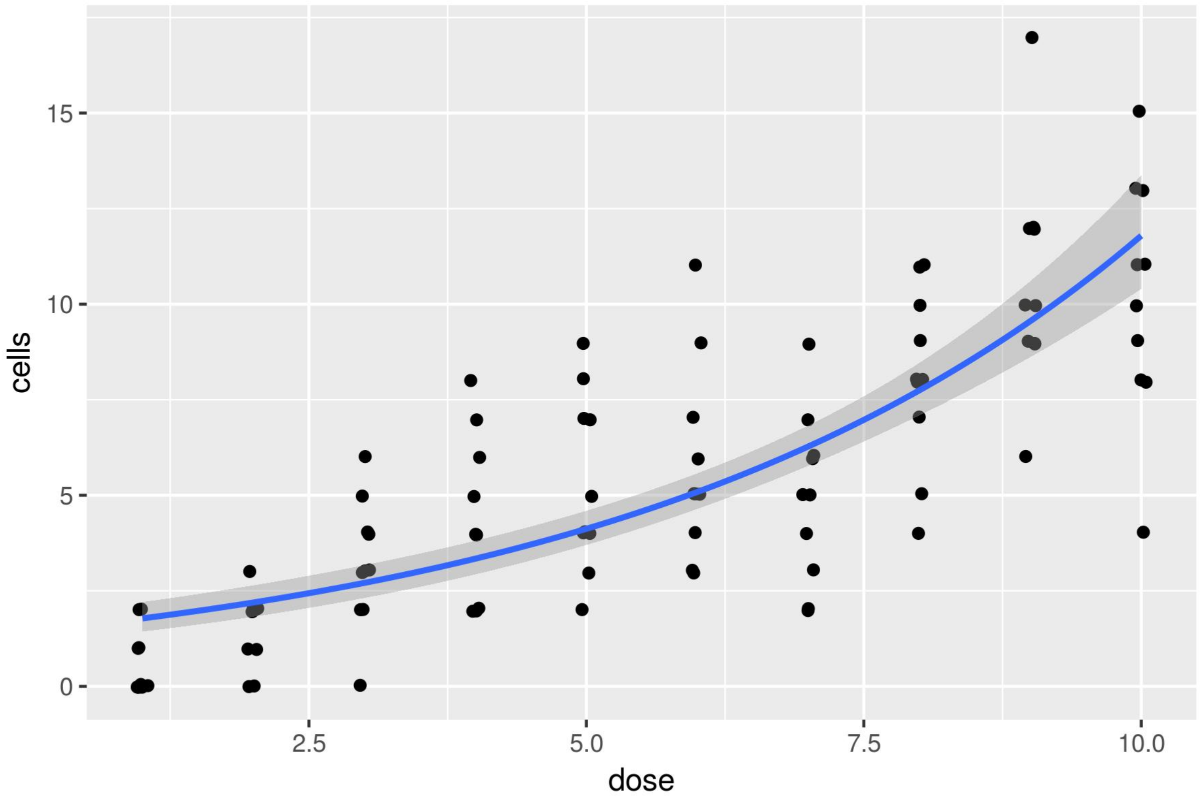
GLMs with geom_smooth()

```
ggplot(data = dat, aes(x = dose, y = cells)) +
   geom_jitter(width = 0.05, height = 0.05)
   geom_smooth(method = 'glm')
```



Poison GLM with geom_smooth()

```
ggplot(data = dat, aes(x = dose, y = cells)) +
   geom_jitter(width = 0.05, height = 0.05) +
   geom_smooth(method = 'glm', method.args = list(family = 'poisson'))
```



Summary of steps

- Plot non-overlapping points
- Add in Poisson trend line
- Polish figure (not-done here)

Let's practice!

GENERALIZED LINEAR MODELS IN R

Understanding output from logistic regression

GENERALIZED LINEAR MODELS IN R

Richard Erickson
Instructor

Communicating results from logistic regression?

- Linear regression is straight forward:
    - Add intercepts
    - Multiply slopes
- Poisson regression:
    - Requires exponential transformation
    - Similar to linear regression post-transformation
- Logistic regression???

Odds-ratios

- Not as straightforward as Poisson exponential
- Used to compare relative odds of two events occurring

Example odds-ratios

- Unfair coin:
  - Compare heads to tails
  - Heads 3 times for every 1 tails
  - 3-to-1 odds
  - Odds-ratios 3.0
- Often used in sports/gambling
- Medical studies

Logistic derivation of odds-ratio

Log-odds ("logit"):

$$\phi(x) = \ln(\tfrac{p(x)}{1-p(x)}) = \beta_0 + \beta_1 x$$

Odds, take exponential (e^x):

$$\tfrac{p(x)}{1-p(x)} = e^{\beta_0 + \beta_1 x}$$

Odd-ratio for continuous variable

Odds-ratio (OR) for continuous variable:

$$\text{OR} = \frac{e^{\beta_0+\beta_1(x+1)}}{e^{\beta_0+\beta_1 x}} = e^{\beta_1}$$

Interpretation

OR Values:

- OR = 1: Coefficient has no effect
- OR < 1: Coefficient decreases odds
- OR > 1: Coefficient increases odds

Cancer example

Non-smoking vs smoking males (Pesch et al. 2012)

- OR: 103.5 (95% CI 74.8-143.2)
- $>$ 100-to-1 odds of getting cancer for smoking men!
- Medical literature often reports 95% confidence intervals rather than p-values
- Broader trend away from p-values

Extract from GLM

```
glm_out <- glm(y ~ x, family = 'binomial')

coef(glm_out)

exp(coef(glm_out))

confint(glm_out)

exp(confint(glm_out))
```

Tidyverse

```
library(broom)

glm_out <- glm(y ~ x, family = 'binomial')

tidy(glm_out, exponentiate = TRUE, conf.int= TRUE)
```

Let's practice!

GENERALIZED LINEAR MODELS IN R

ggplot2 and binomial GLM

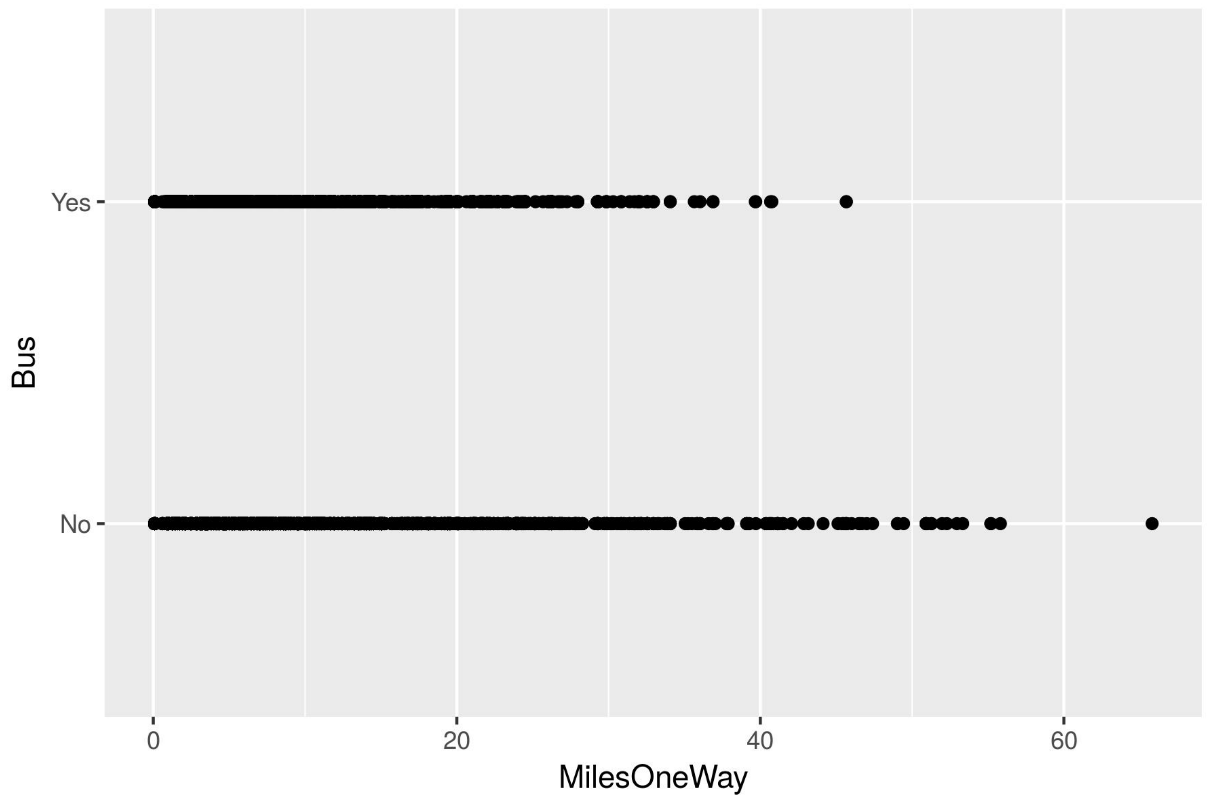
GENERALIZED LINEAR MODELS IN R

Richard Erickson
Instructor

What can I see in my data?

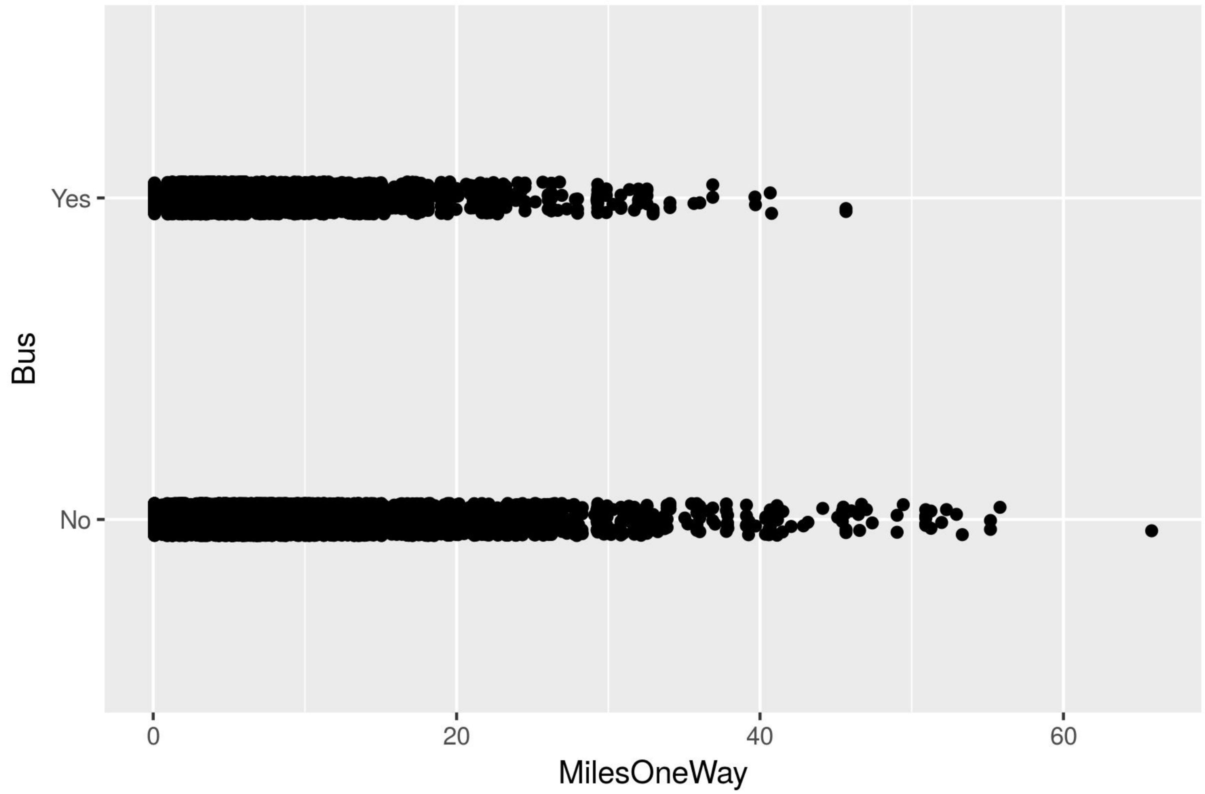
Does commute distance change the probability of taking the bus?

```
ggplot(bus, aes(x = MilesOneWay, y = Bus)) + geom_point()
```



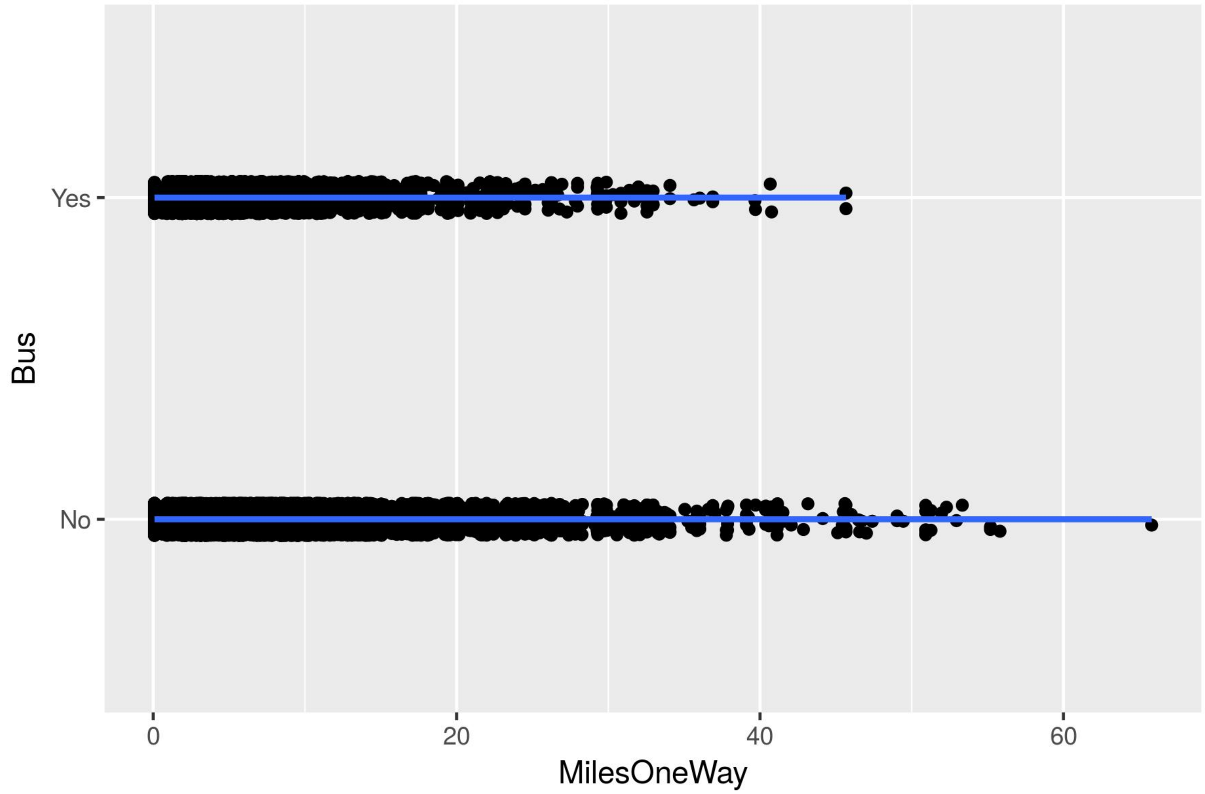
geom_jitter()

```
gg_jitter <- ggplot(bus, aes(x = MilesOneWay, y = Bus)) +
             geom_jitter(width = 0, height = 0.05)
print(gg_jitter)
```



geom_smooth()

```
gg_jitter + geom_smooth()
```

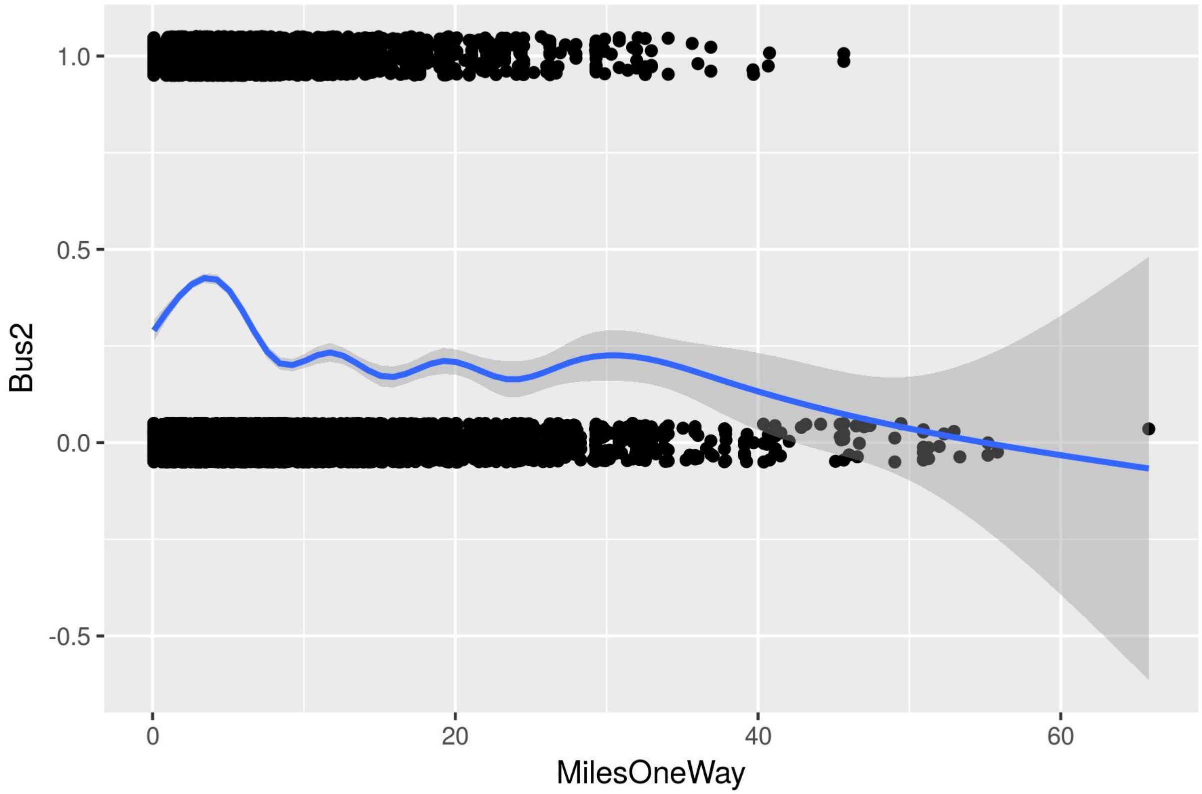


factor to numeric

```
str(bus)
bus$Bus2 <- as.numeric(bus$Bus) - 1
```

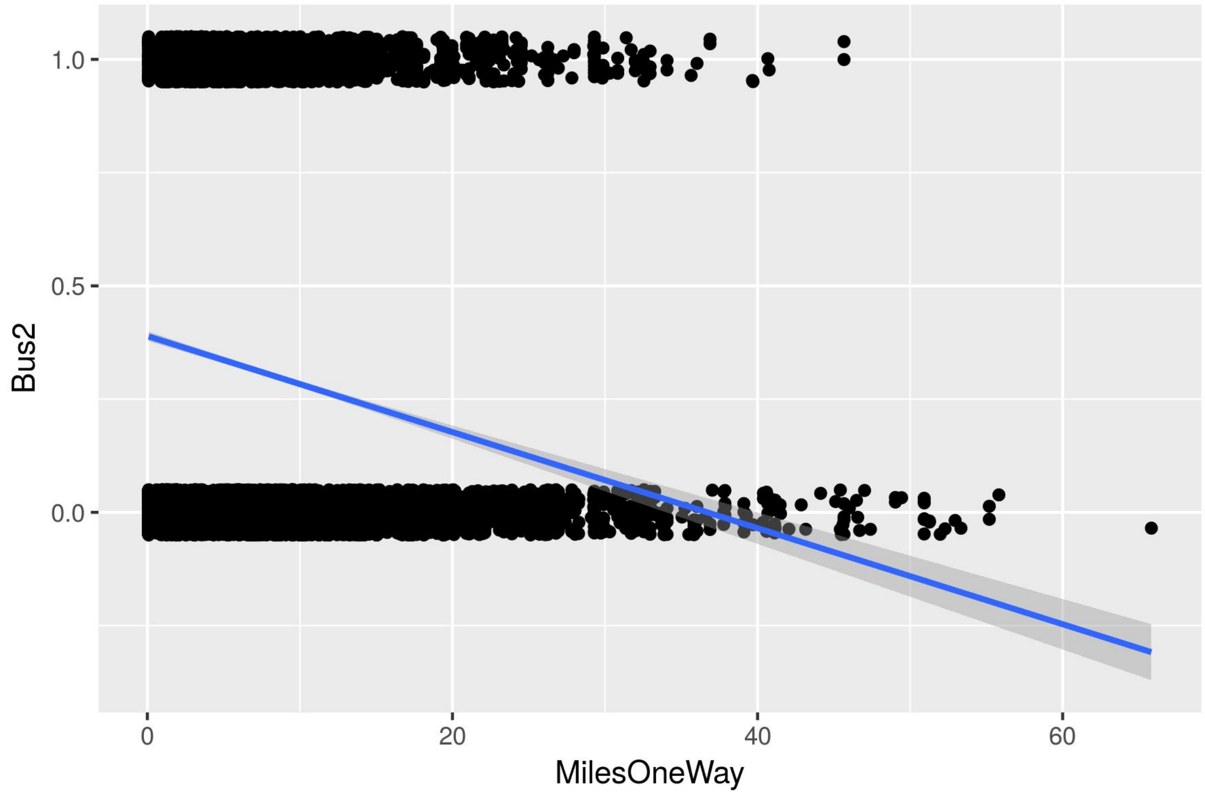
geom_smooth()

```
gg_jitter + geom_smooth()
```



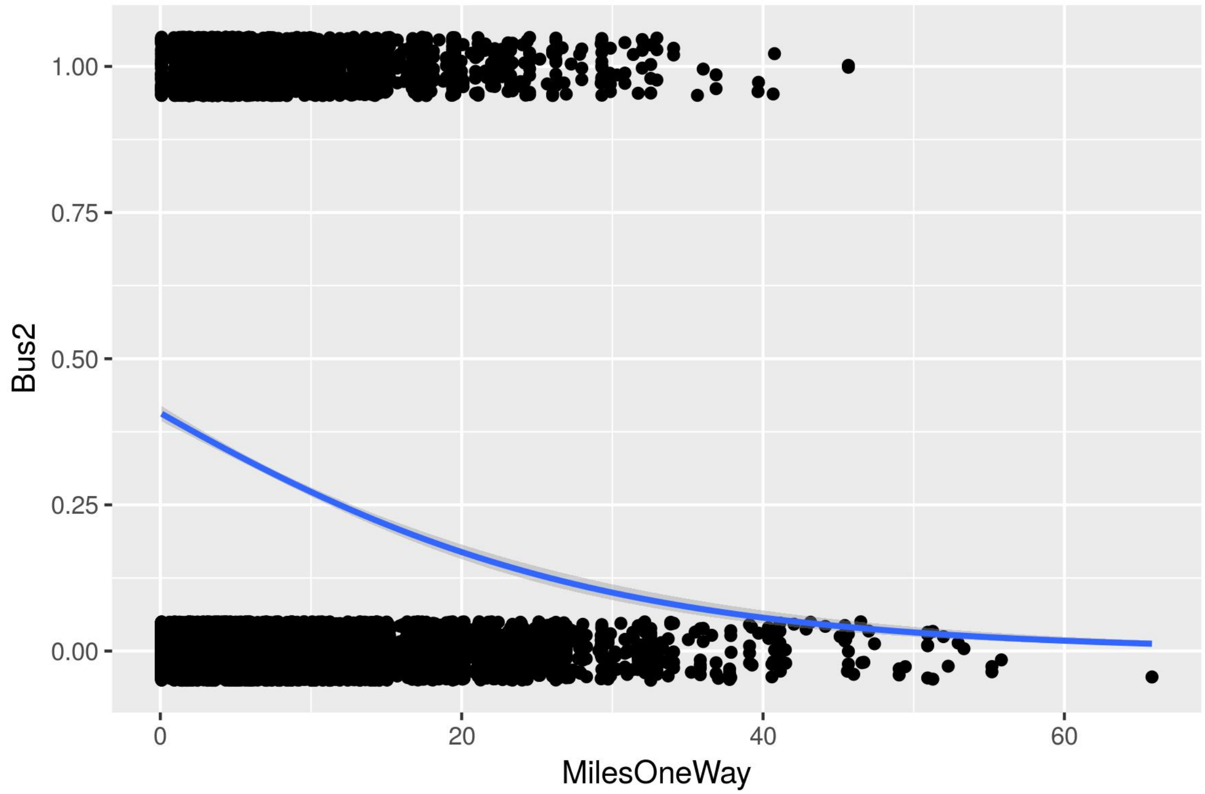
linear models

```
gg_jitter + geom_smooth(method = 'glm')
```

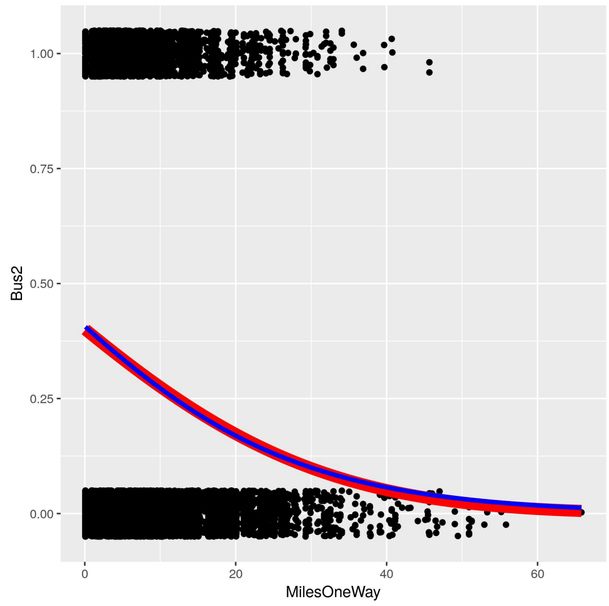


Logistic regressions

```
ggJitter +
geom_smooth(method = 'glm',
            method.args = list(family = "binomial"))
```



```
gg_jitter +
    geom_smooth(method = 'glm',
                method.args = list(family = binomial(link = 'logit')),
                se = FALSE, color = 'red') +
    geom_smooth(method = 'glm',
                method.args = list(family = binomial(link = 'probit')),
                se = FALSE, color = 'blue')
```



Summary of steps

- Plot as jitter to avoid overlap
- Add a smoothed geom
- Specify correct method and family
- Polish your figure (not covered in this course)

Let's practice!

GENERALIZED LINEAR MODELS IN R