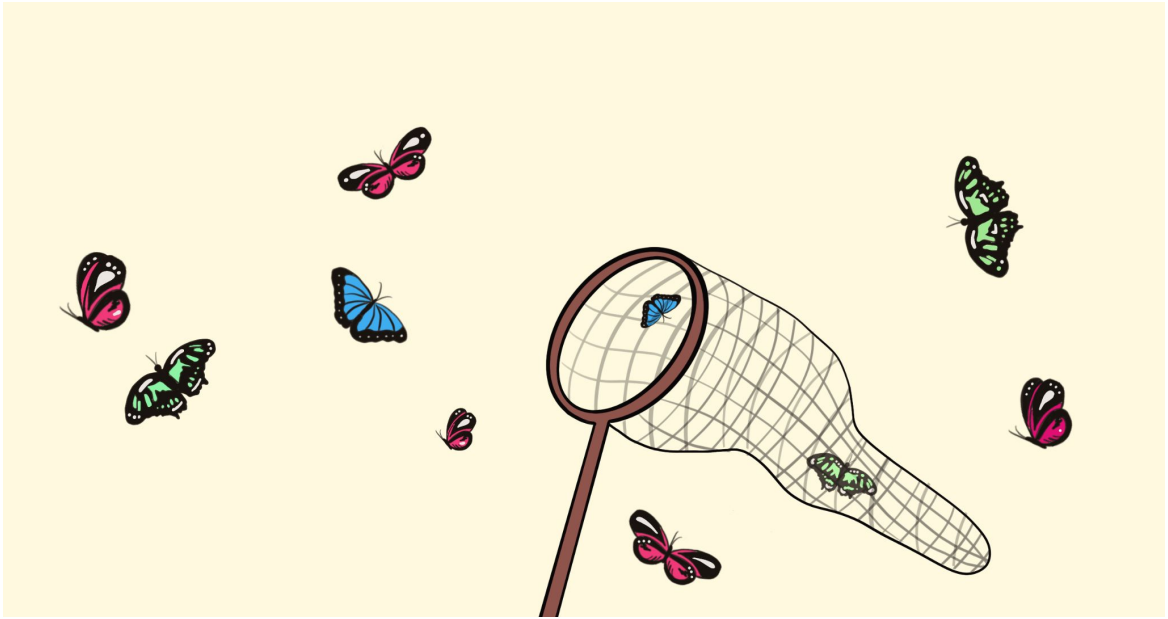
Week 8 Lecture 1:
Poisson regression

EDS 222: *Statistics for Environmental Data Science*

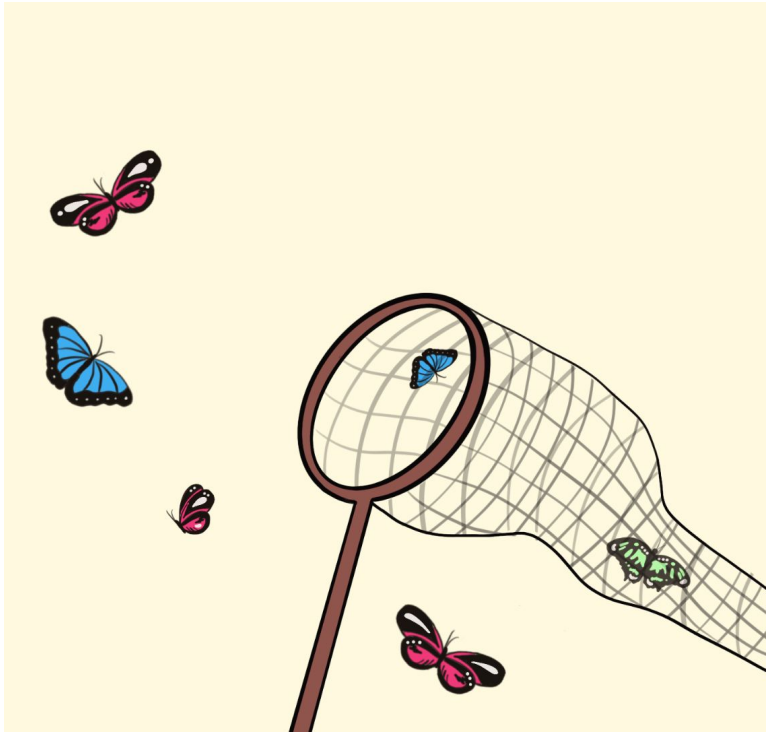


California wildfires

Jessica Christian / The Chronicle

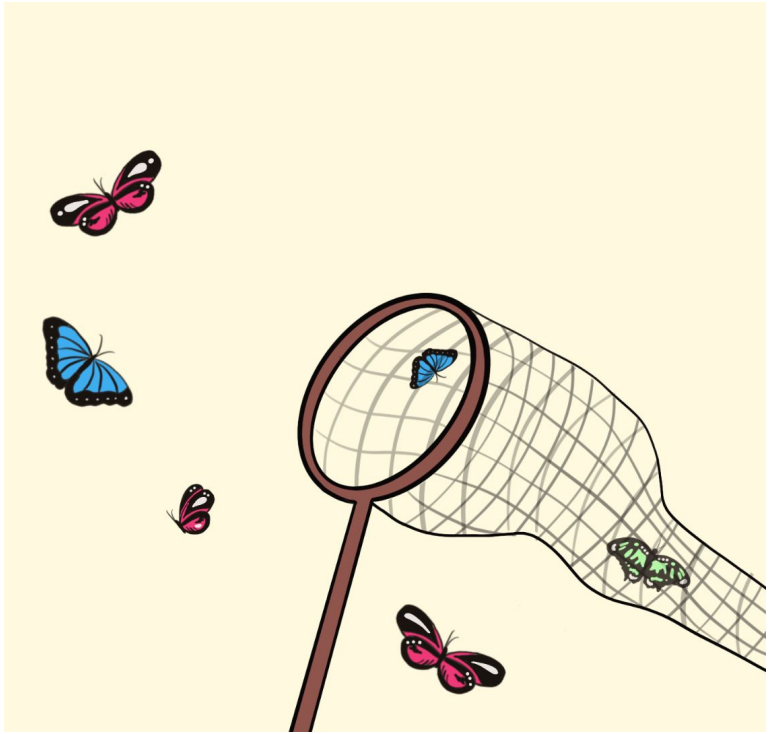
Today’s agenda

- ➔ Poisson variables
- ➔ Poisson regression
- ➔ Begin on the lab

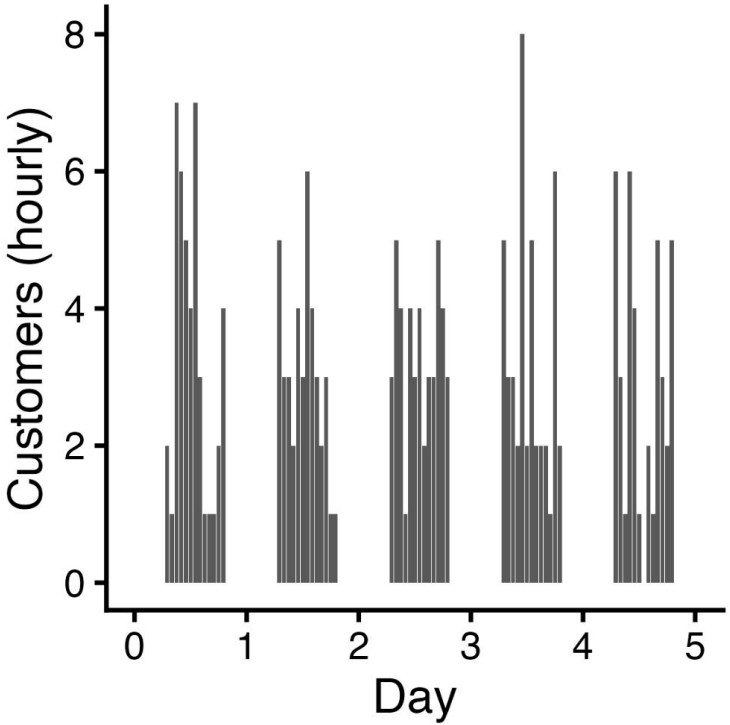


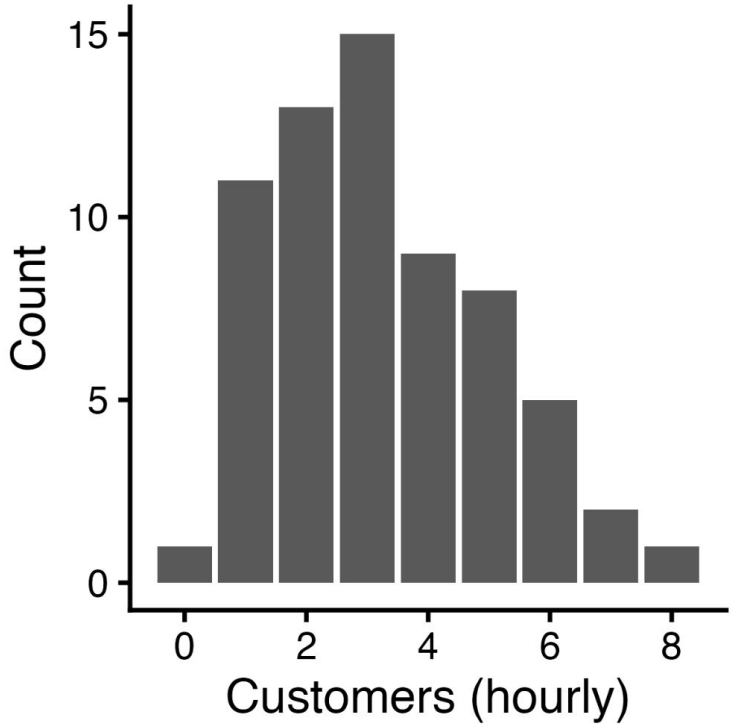
Today’s agenda

- ➔ **Poisson variables**
- ➔ Poisson regression
- ➔ Begin on the lab

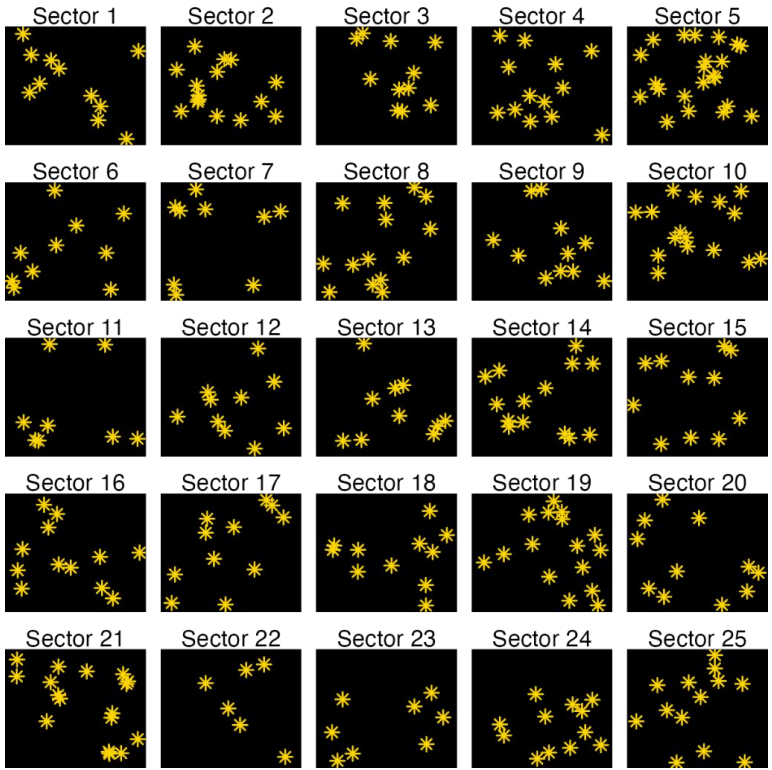


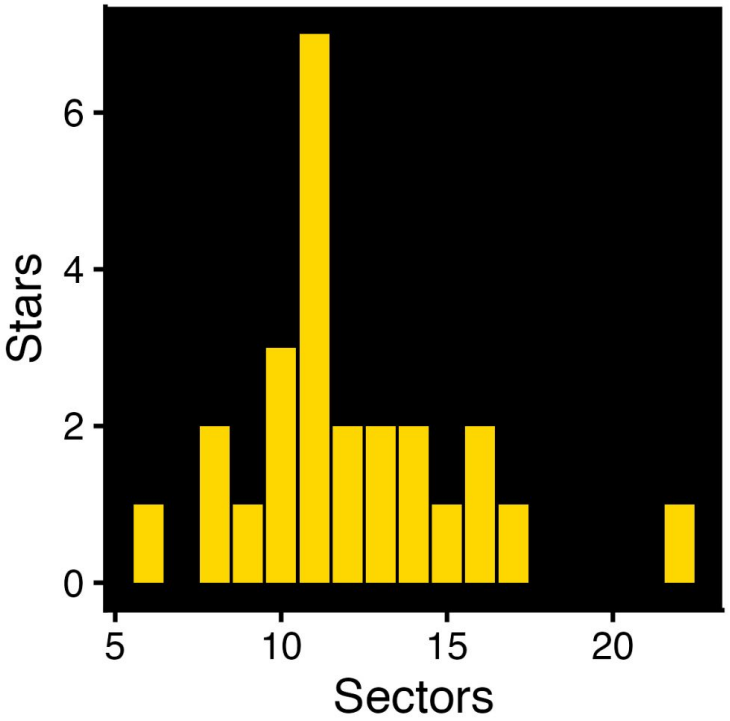
Examples of Poisson variables





Examples of Poisson variables

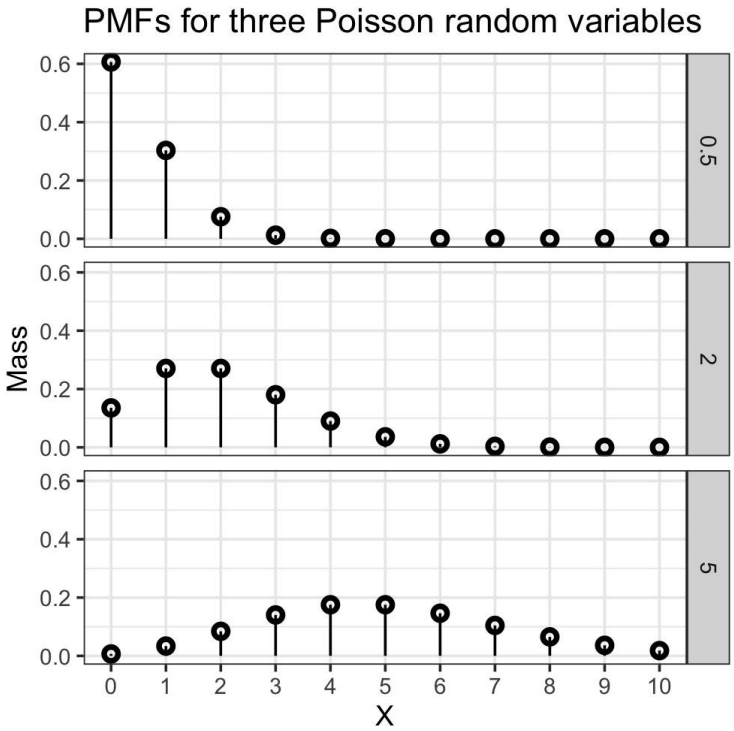




Characteristics of Poisson variables

- A fixed window. Time, space … anything else that's a fixed "width" (e.g., mutations on a chromosome).
- Independent events happening at a fixed rate. Customers arriving, stars being born, species observed on a transect.
- One parameter: λ.
  - The expected value of Poisson(λ) is λ.
  - The standard deviation of Poisson(λ) is sqrt(λ).
  - λ controls BOTH mean and variance.
- If λ varies across windows (e.g., events happen in clusters), Poisson is too restrictive. Use negative binomial instead. Option for the final project!

Poisson PMFs



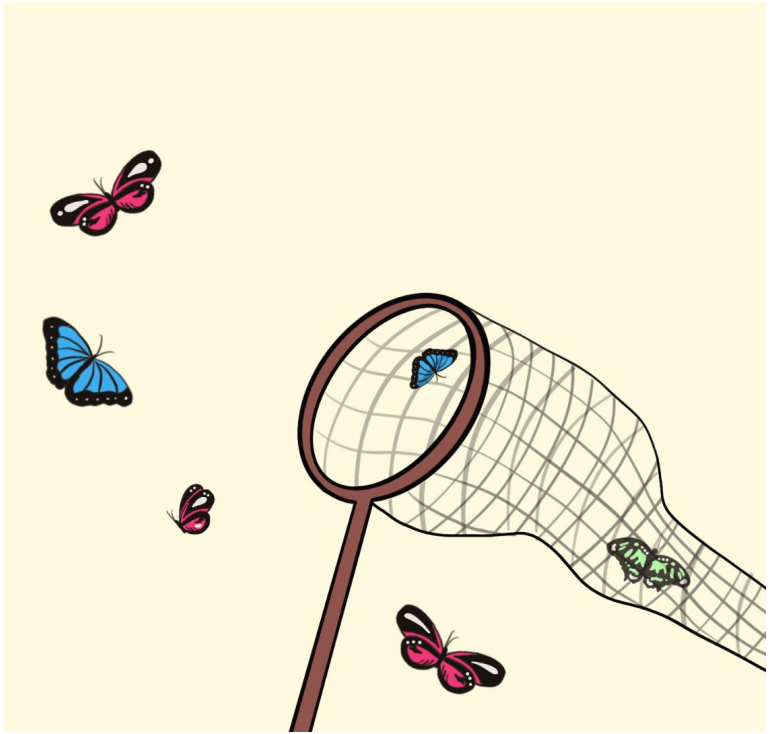
You created these PMFs before class

Think about what each might represent. What events could be happening? What's the "size" of the window?

Take a minute to think on your own, then discuss with a peer.

Today’s agenda

- ➔ Poisson variables
- ➔ **Poisson regression**
- ➔ Begin on the lab

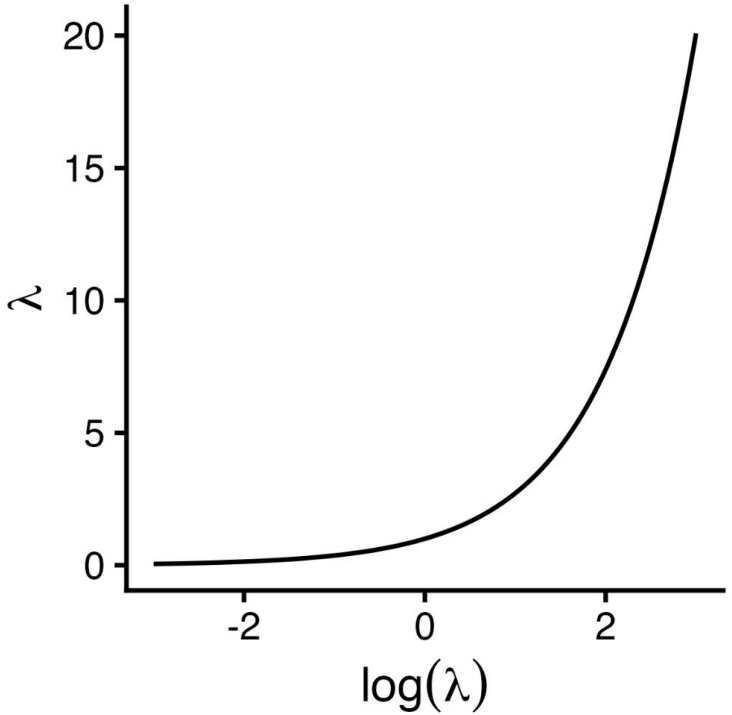


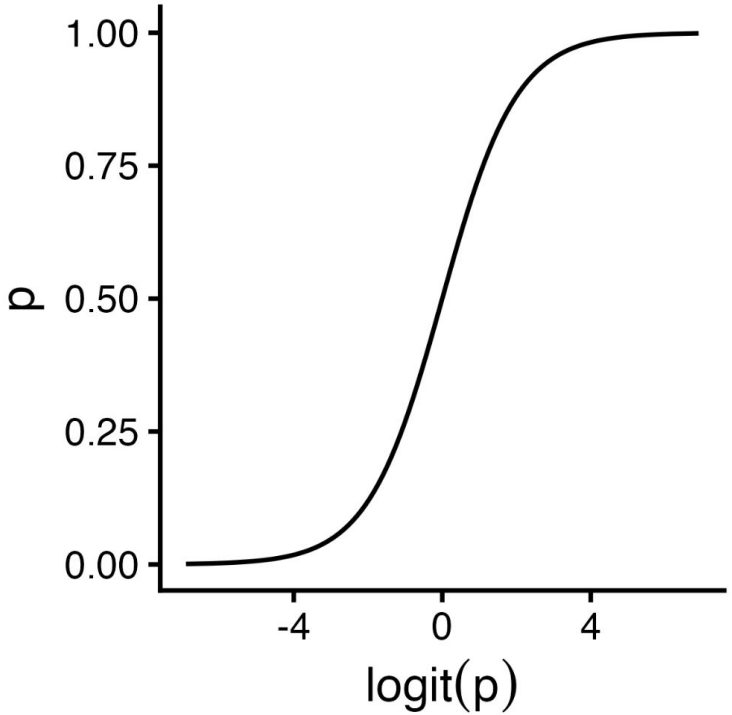
Poisson statistical notation

$$\text{CountOutcome} \sim Poisson(\lambda)$$

$$log(\lambda) = \beta_0 + \beta_1 \text{Predictor}$$

Comparing link functions





Wildfire occurrence

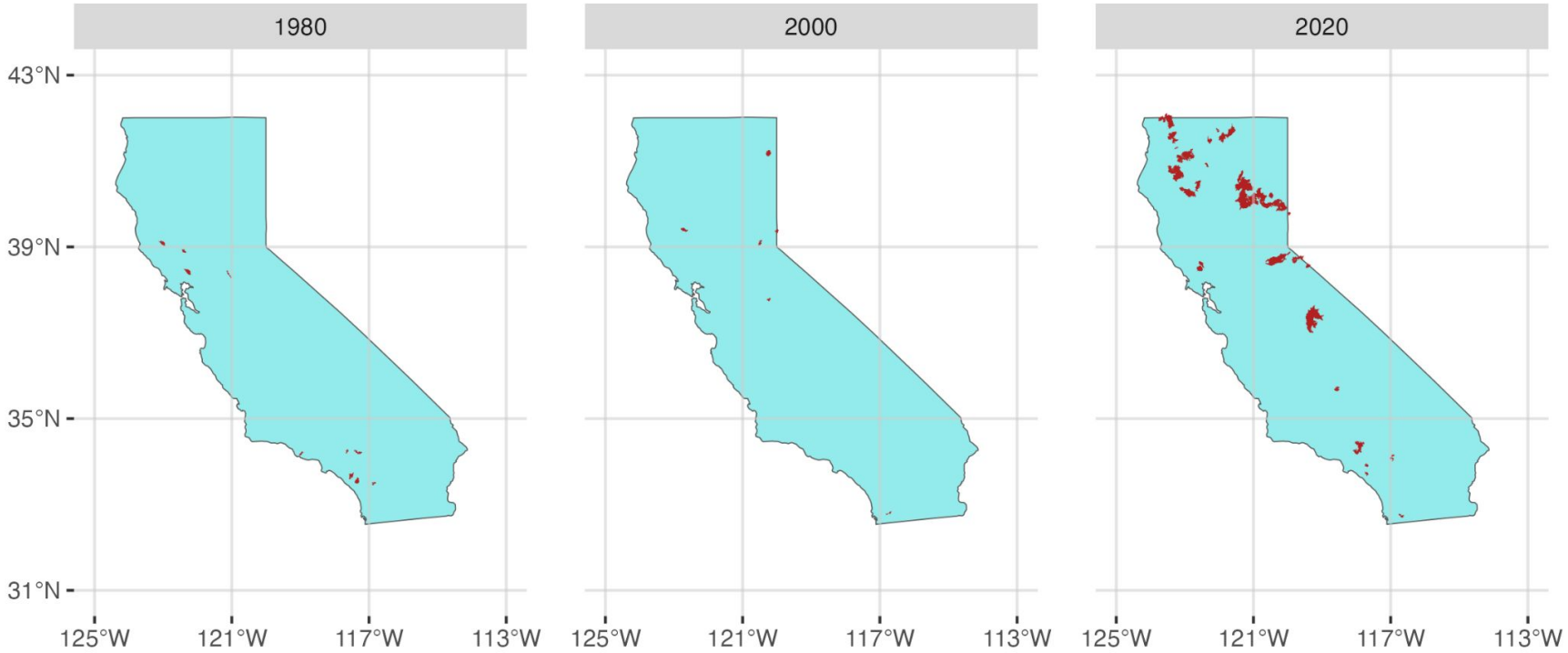
Jessica Christian / The Chronicle

Model notation

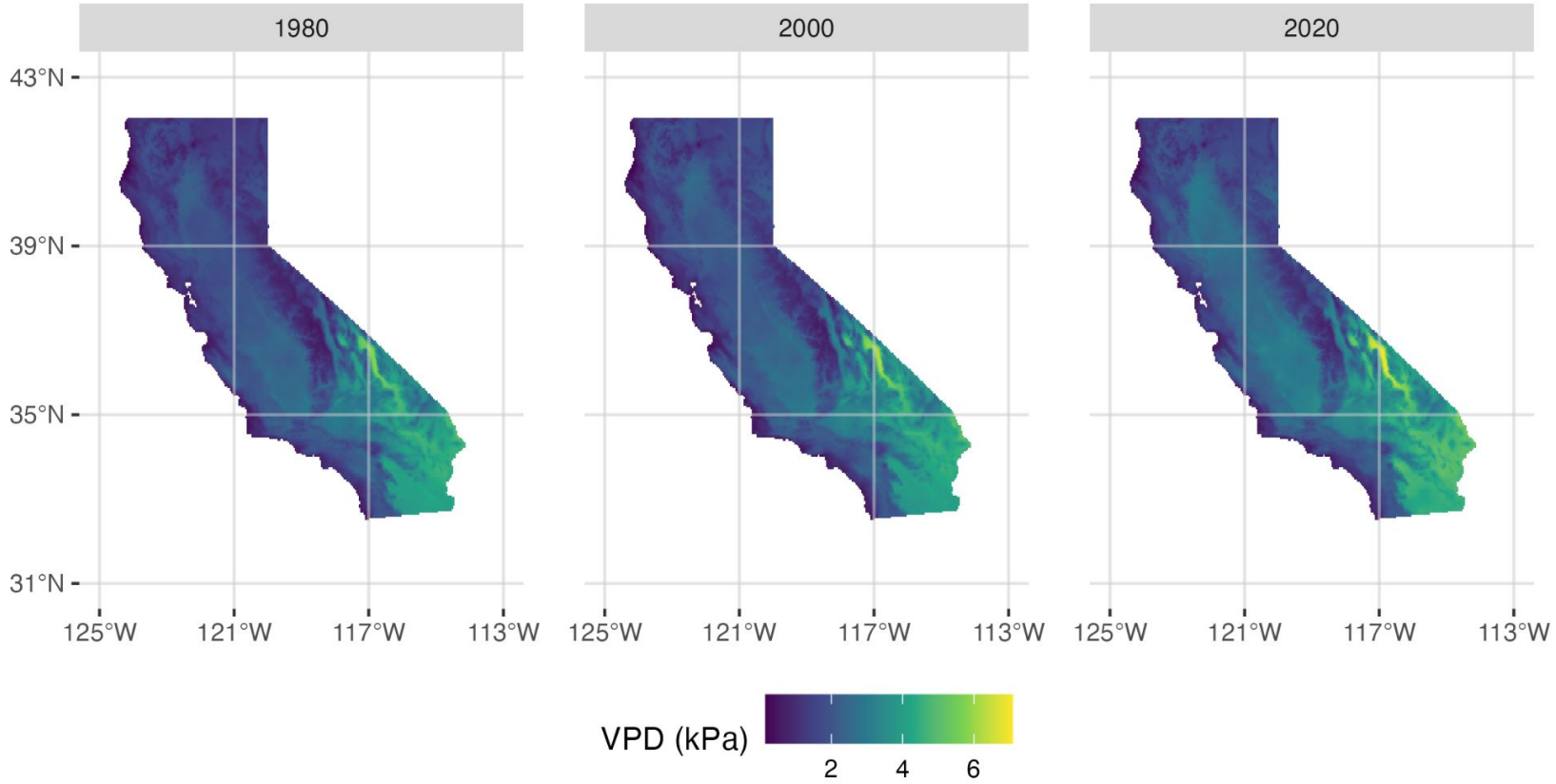
$$\text{Wildfires} \sim Poisson(\lambda)$$

$$log(\lambda) = \beta_0 + \beta_1 \text{Summer VPD}$$

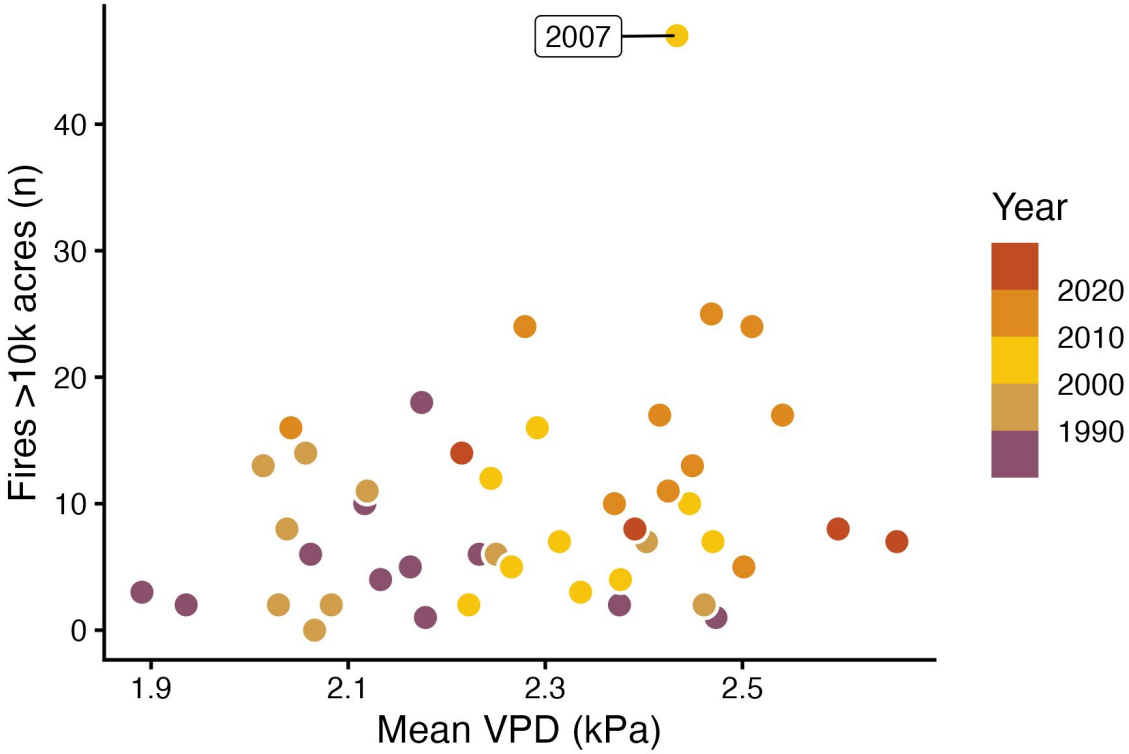
Available data



Available data



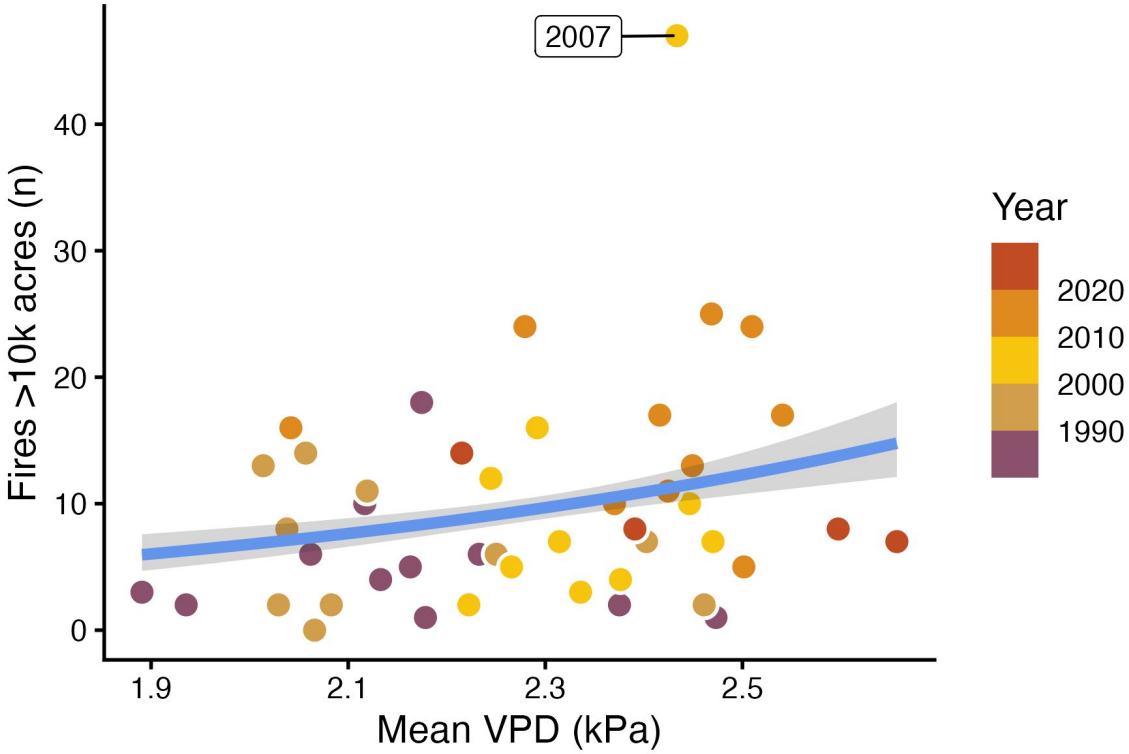
Available data



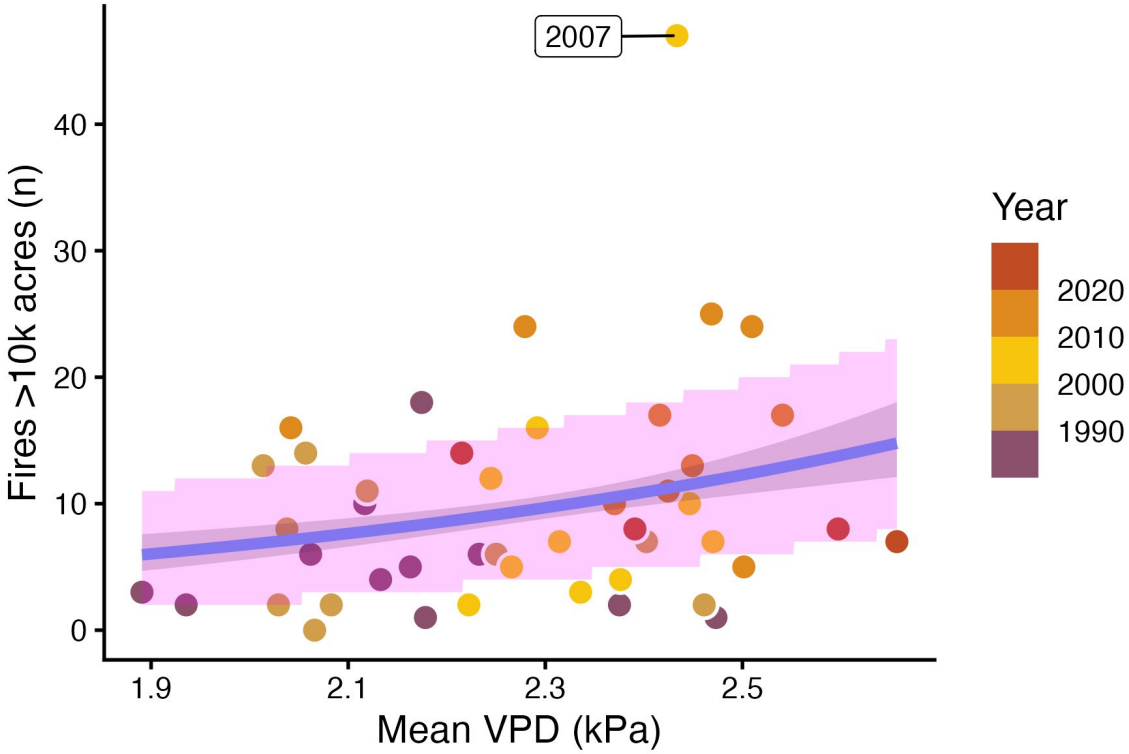
Model summary

```
> summary(fire_mod_pois)

Call:
glm(formula = n_fires ~ mean_vpd_kpa, family =
poisson(link = "log"),
    data = wildfire_weather)

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)   -0.4550     0.6138  -0.741    0.458
mean_vpd_kpa   1.1851     0.2640   4.490 7.13e-06 ***
```

Visualized predictions



Overdispersion



Recap

- Poisson variables describe counts.
  - One parameter, λ. The mean of the distribution is λ.
- Poisson regression uses log() for the link function.
- Predictions, CIs, and p-values should be interpreted as with other GLMs (e.g., logistic).
  - For distributions of the response variable, mean ± 2SD no longer applies. Use quantiles (`qpois()`) instead.
- λ controls both mean AND variance, so Poisson is often too restrictive for real data.
  - Negative binomial is often a better alternative and an option for your final project.

Today’s agenda

- ➔ Poisson variables
- ➔ Poisson regression
- ➔ **Begin on the lab**