



ECON 2041

Introductory Econometrics

Untangling relationships

# Plan for the semester

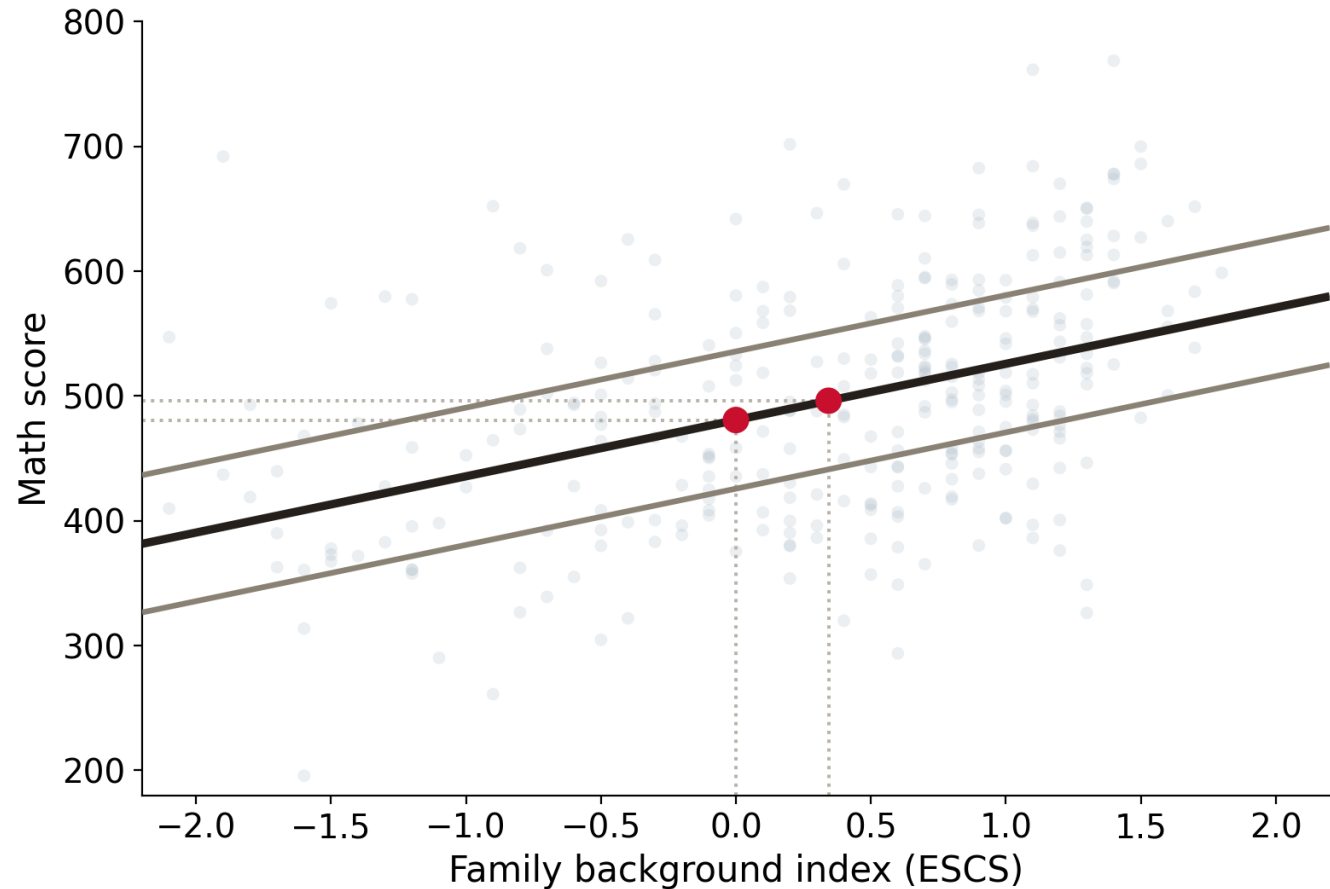


# Plan for today

- Finish the line: how we pick & interpret the intercept
- Name the parts of a regression
- Dataset vote: last call
- How lucky is the Lucky Country?
- Your intuition vs. the data
- More than one X: multivariate regression

# The intercept anchors the line at the means

$$\begin{aligned}\bar{y} &= \hat{\beta}_0 + \hat{\beta}_1 \bar{x} \\ \hat{\beta}_0 &= \bar{y} - \hat{\beta}_1 \bar{x} \\ &= 496.2 - 45.07(0.343) \\ &= 480.7\end{aligned}$$



# Interpretation

$$\widehat{\text{math}} = 481 + 45 \times \text{ESCS}$$

**Slope:** students whose ESCS index is 1 point higher score 45 points higher on average

**Intercept:** a student at ESCS = 0 is predicted to score 481

**Here 0 is meaningful:**

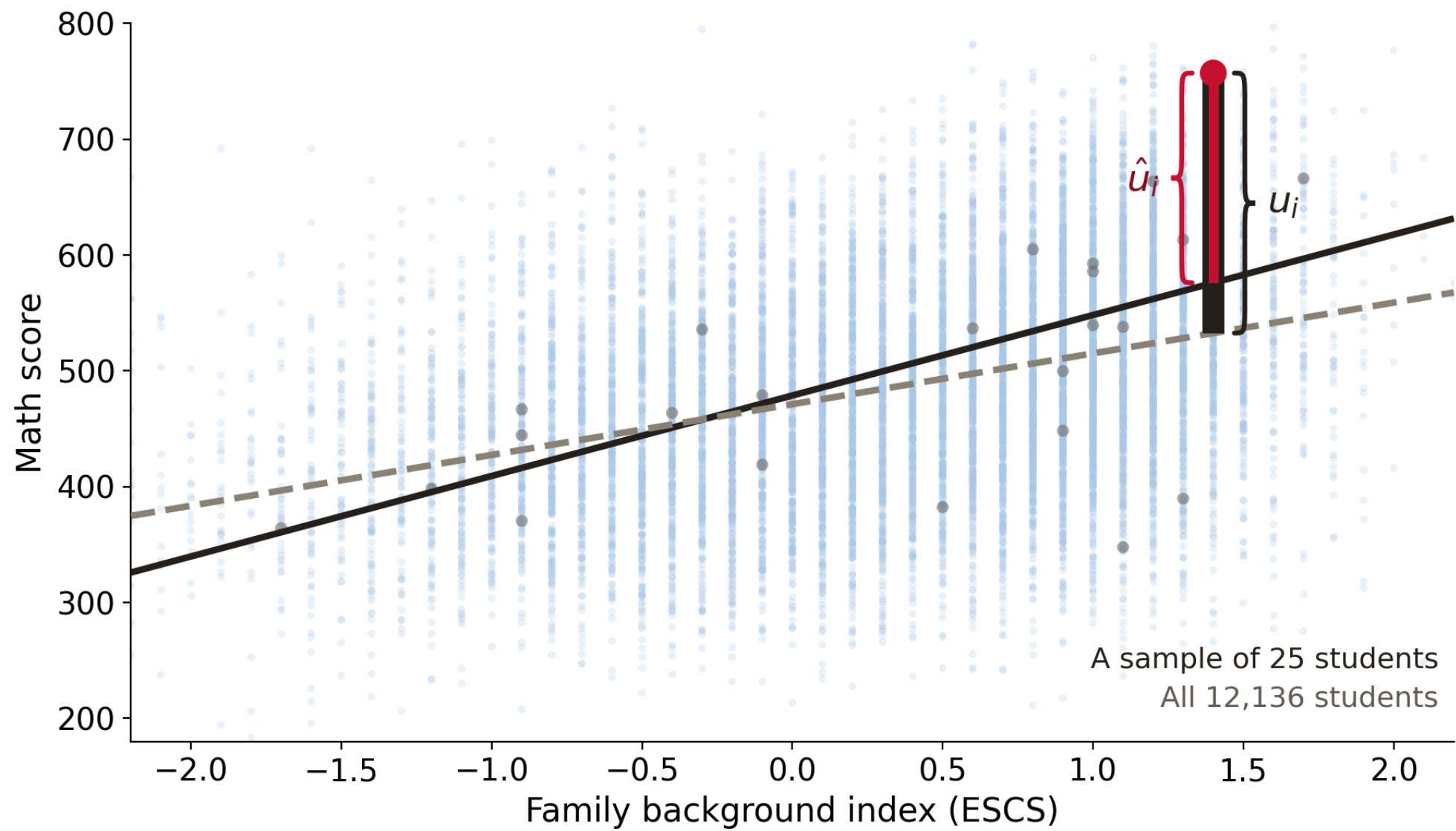
- ESCS = 0 is an OECD-average family
- 481 describes a real kind of student

**Other times it is meaningless:**

- if we regress weight on height
- the intercept is the predicted weight at height 0

Read the intercept, then decide whether it's meaningful

# Residual vs. error (N=25)



Data: PISA 2022, Australia (OECD); dashed line fitted on all 12,136 students, solid line on a sample of 25

# Single regression summary

- Covariance has nonsense units, but we can rescale it to get
  - the correlation coefficient
  - our regression slope
- A straight line describes  $E[Y \mid X]$  when the conditional means line up
- Least squares picks the line with the smallest sum of squared residuals
- Residuals are signed gaps to the fitted line
- Errors are signed gaps to the population line

# The many names of Y and X

$$\widehat{\text{math}} = 481 + 45 \times \text{ESCS}$$

$$y_i = \beta_0 + \beta_1 x_i + u_i$$

$Y$ : the thing we're trying to understand

- **dependent variable**
- **outcome variable**
- response variable
- predicted variable
- LHS variable

$X$ : the variable(s) we use to "explain"  $Y$

- **independent variable**
- **explanatory variable**
- regressor
- predictor
- covariate
- RHS variable

You pick the data!

# Choose your own adventure: last call

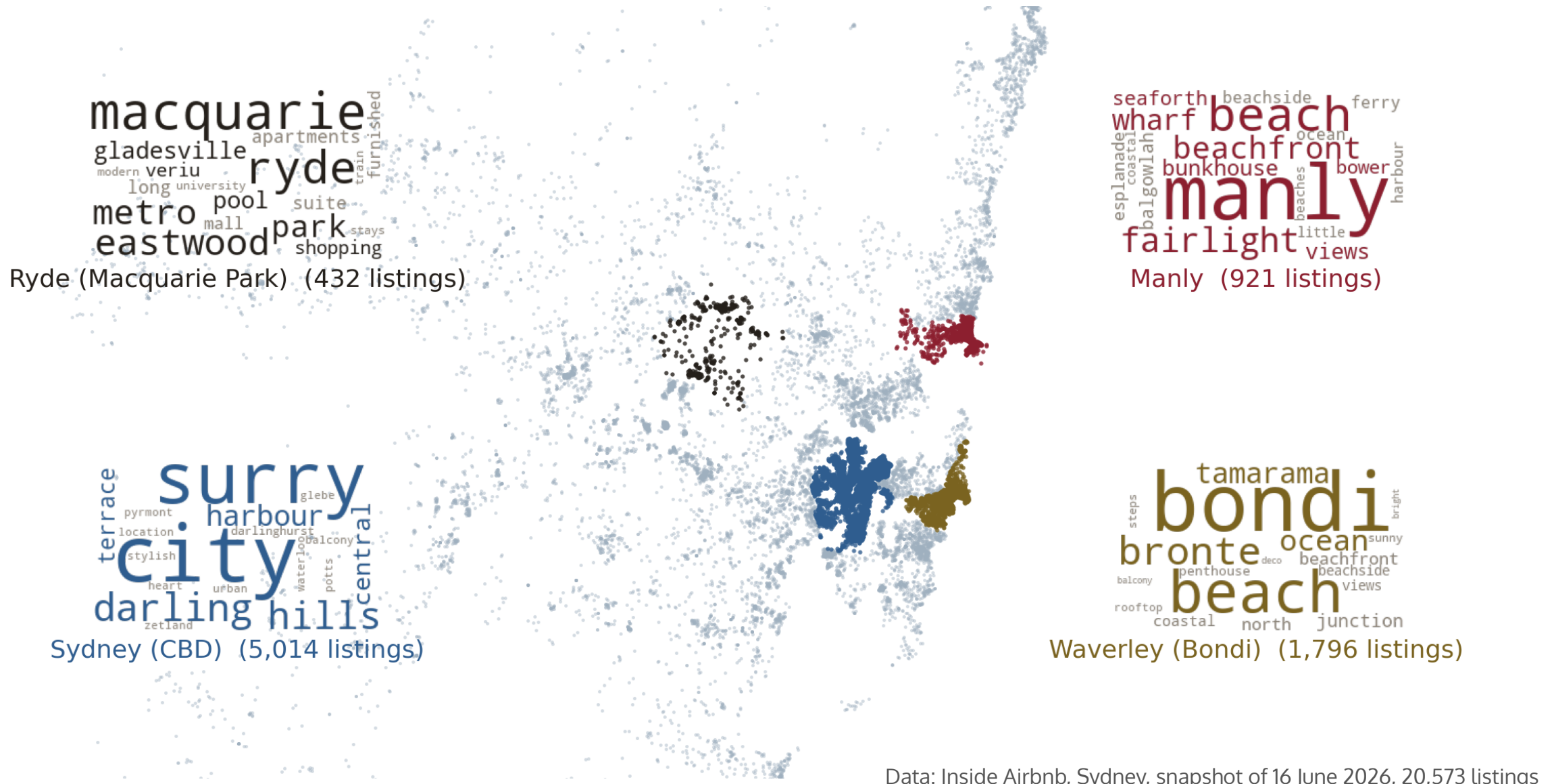
This is your chance to decide which data we work with:

- **Sydney Airbnb: what makes a listing expensive?**  
about 20,000 Sydney listings
- **US wages: who earns what, and why?**  
nearly 150,000 workers

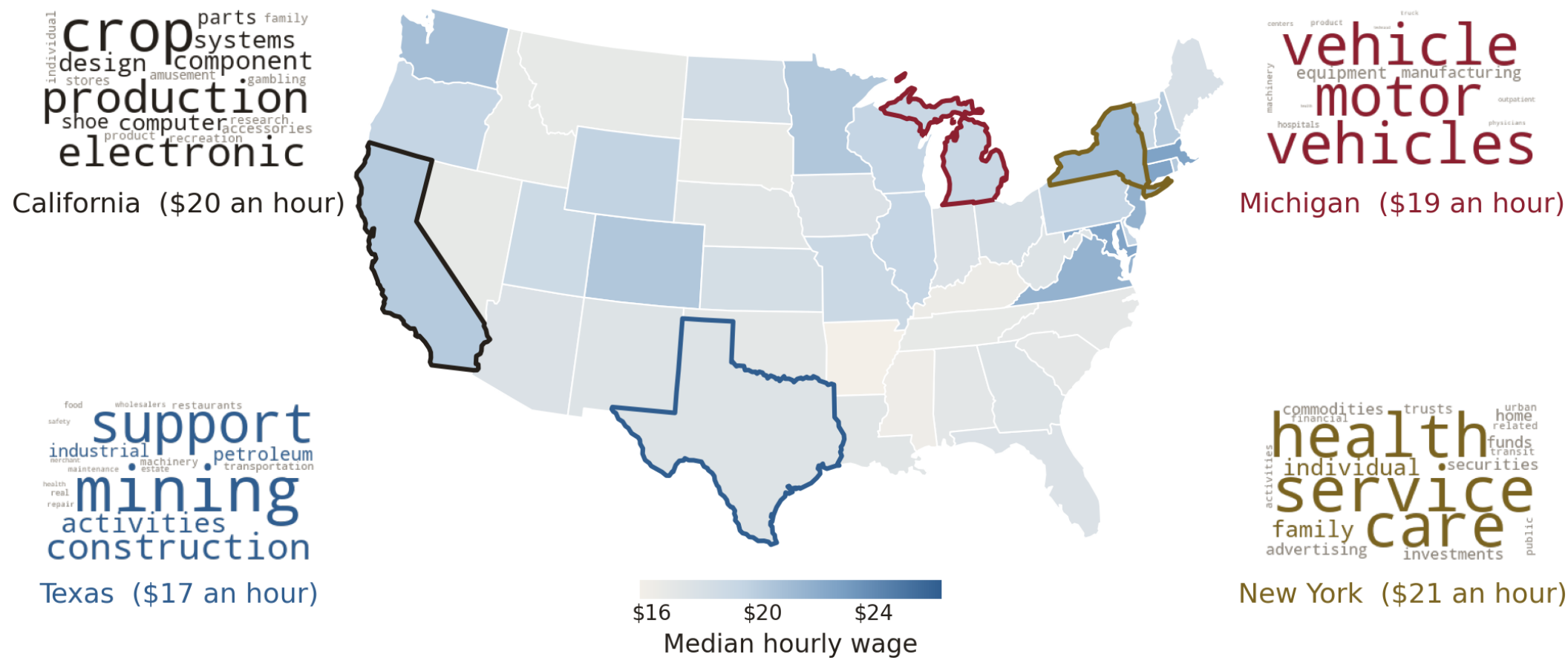
<https://emiliatjernstrom.com/econ2041/portal.html>



# Bondi sells beach, Macquarie Park sells... metro



# Different jobs & different pay



Data: US Current Population Survey, 2014 outgoing rotation groups, 131,082 full-time workers

# How lucky is the Lucky Country?

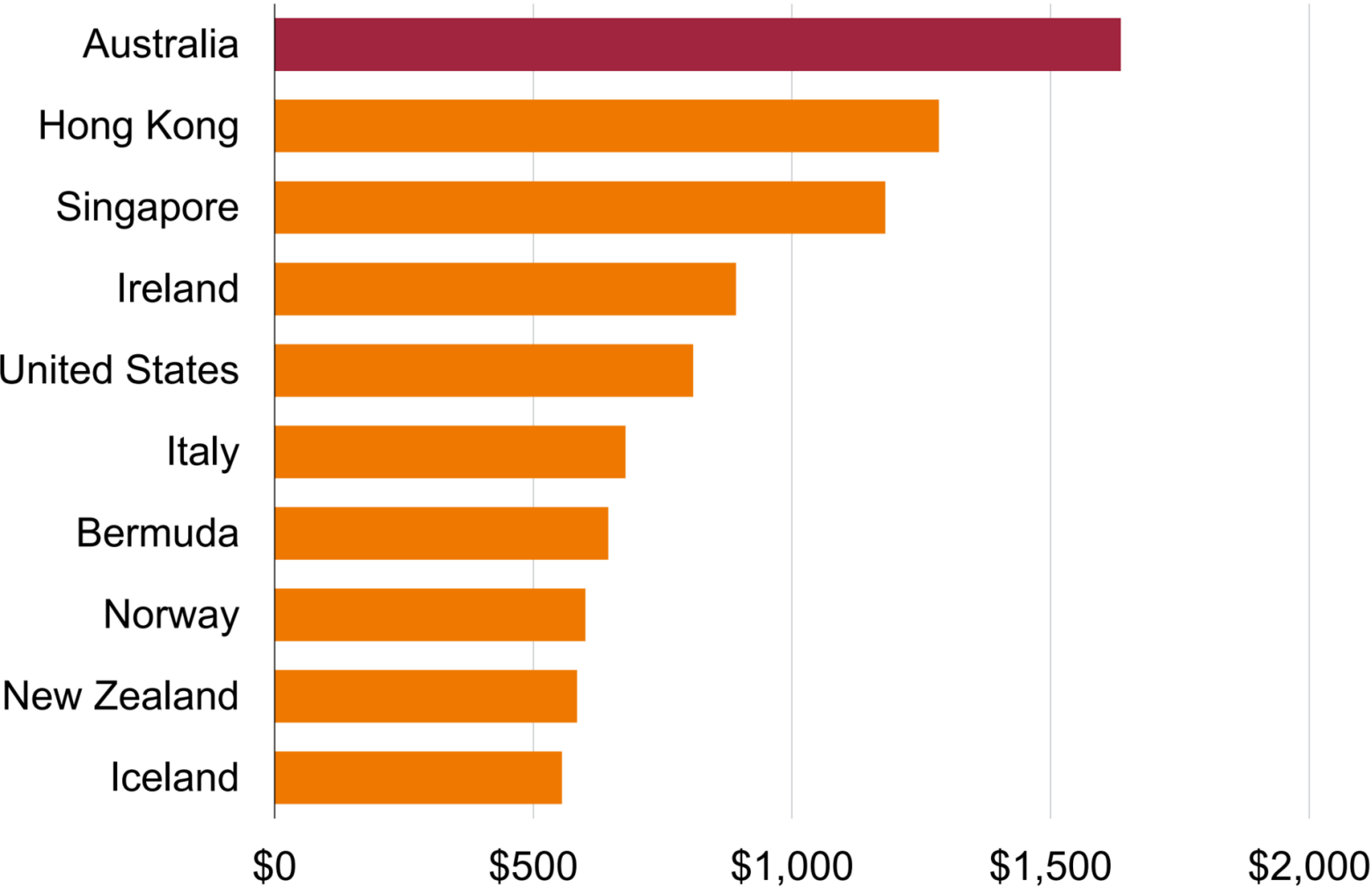
Today's data: pokies and losses across Victoria

# Australians are world leaders...



# Australians are world leaders...

Gambling losses per adult in 2022, Australian dollars



Data: Grattan Institute, "A better bet" (2024), Figure 1.1

# Australians lost \$24B in a year

## **The scale:**

Collectively, Australians lost \$24B gambling in 2020-21

## **The concentration:**

76% of the world's pub-and-club pokies are in Australia

## **The inequality:**

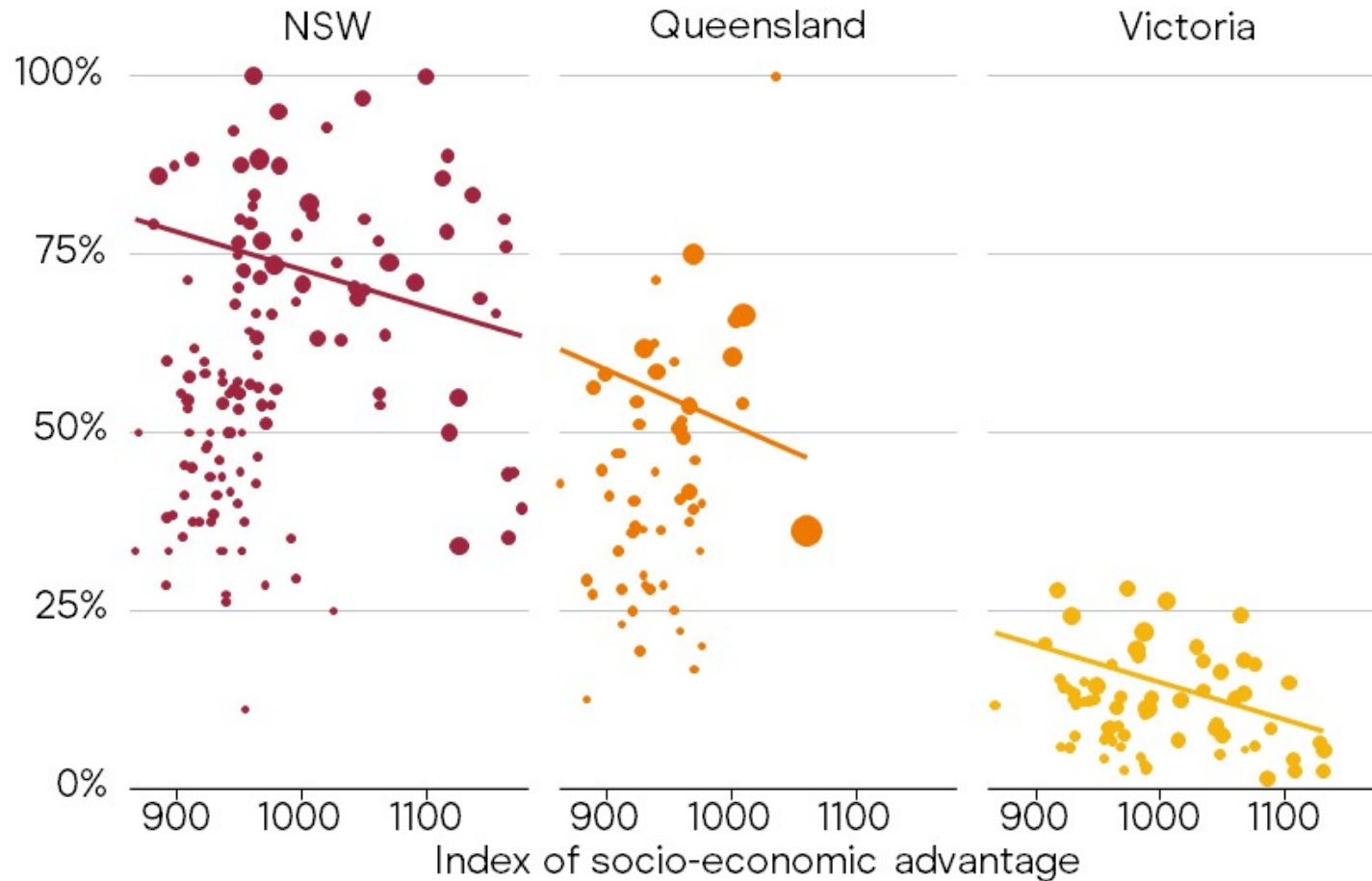
Some communities lose 10x more per capita than others

## **The politics:**

\$9B in state gambling taxes, 7.8% of all state tax revenue

# Pokies can be hard to escape!

Proportion of pubs / clubs in LGA that have pokies



# This is highly policy-relevant & studied

## Electronic gaming machine accessibility and gambling problems: A natural policy experiment

[Alex MT Russell](#)<sup>1,\*</sup>, [Matthew Browne](#)<sup>2</sup>, [Nerilee Hing](#)<sup>2</sup>, [Matthew Rockloff](#)<sup>2</sup>, [Philip Newall](#)<sup>1</sup>, [Nicki A Dowling](#)<sup>3</sup>,  
[Stephanie Merkouris](#)<sup>3</sup>, [Daniel L King](#)<sup>4</sup>, [Matthew Stevens](#)<sup>5</sup>, [Anne H Salonen](#)<sup>6</sup>, [Helen Breen](#)<sup>7</sup>, [Nancy Greer](#)<sup>8</sup>,  
[Hannah B Thorne](#)<sup>9</sup>, [Tess Visintin](#)<sup>9</sup>, [Vijay Rawat](#)<sup>8</sup>, [Linda Woo](#)<sup>10</sup>

► [Author information](#) ► [Article notes](#) ► [Copyright and License information](#)

PMCID: PMC10562817 PMID: [37594879](#)

# Net expenditure means losses

**EGM** (electronic gaming machine): pokies

**LGA:** Local Government Area

**Net expenditure:** Gambling losses

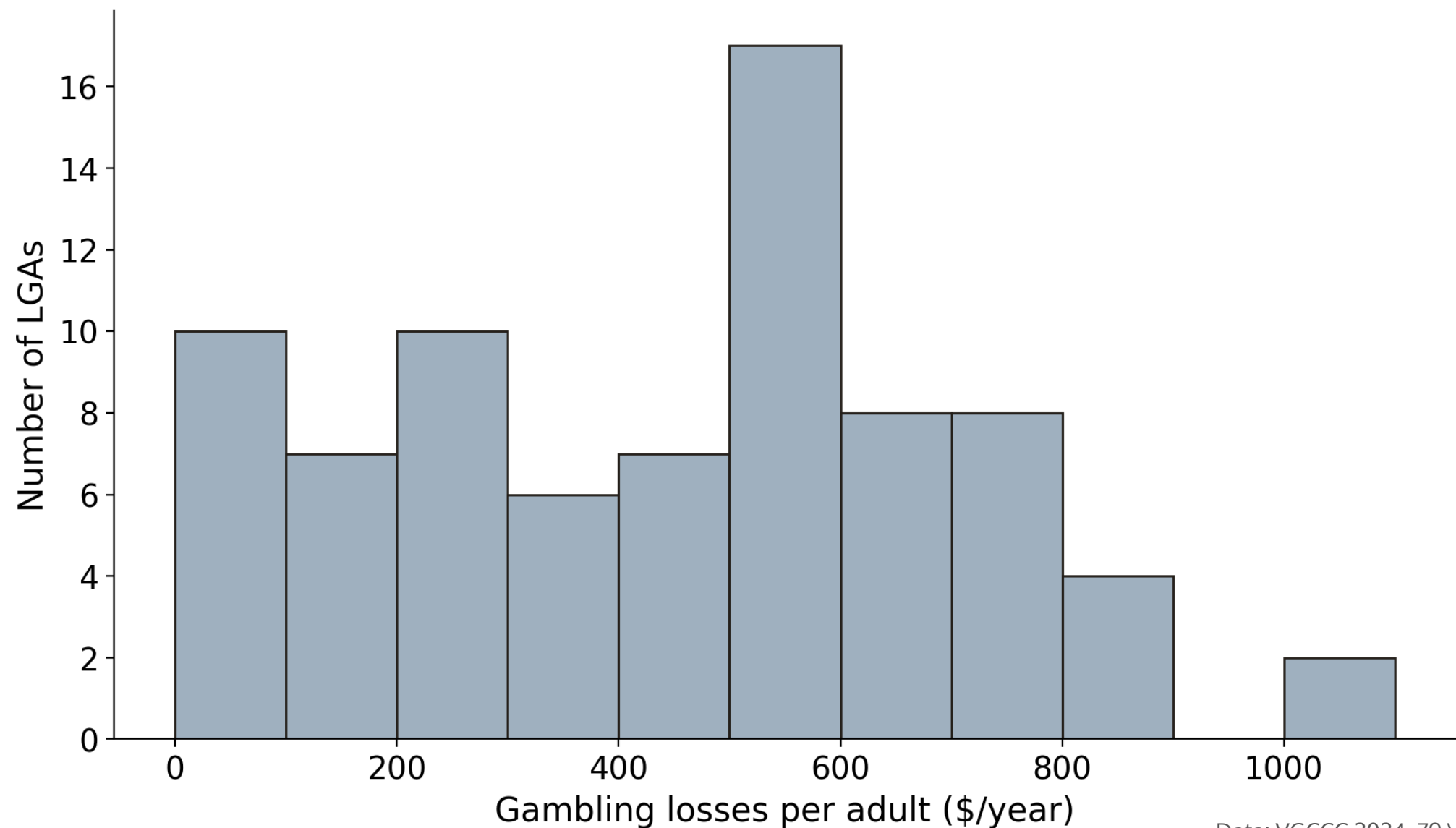
**Data sources:**

- Gaming expenditure: Victorian Gambling and Casino Control Commission
- Employment: Department of Employment
- Population: Department of Environment, Land, Water and Planning

# Seven variables, one row per LGA

|                                                                                      |                                                        |
|--------------------------------------------------------------------------------------|--------------------------------------------------------|
| <code>total_net_exp</code>                                                           | Total net gaming expenditure                           |
| <code>adult_pop</code>                                                               | LGA adult population                                   |
| <code>egm_per_1000</code>                                                            | EGMs per 1,000 adults                                  |
| <code>exp_per_adult</code>                                                           | Expenditure per adult (\$)                             |
| <code>unemployed</code>                                                              | Unemployed persons (count)                             |
| <code>ue_rate</code>                                                                 | LGA unemployment rate                                  |
| <code>SEIFA_dis_score</code>                                                         | SEIFA disadvantage index (higher = less disadvantaged) |
| <code>total_net_exp</code> and <code>exp_per_adult</code> are <b>gambling losses</b> |                                                        |

# Losses per adult vary tenfold across LGAs



Data: VGCCC 2024, 79 Victorian LGAs

# Let's check your intuition!

3 predictions, then the data answer

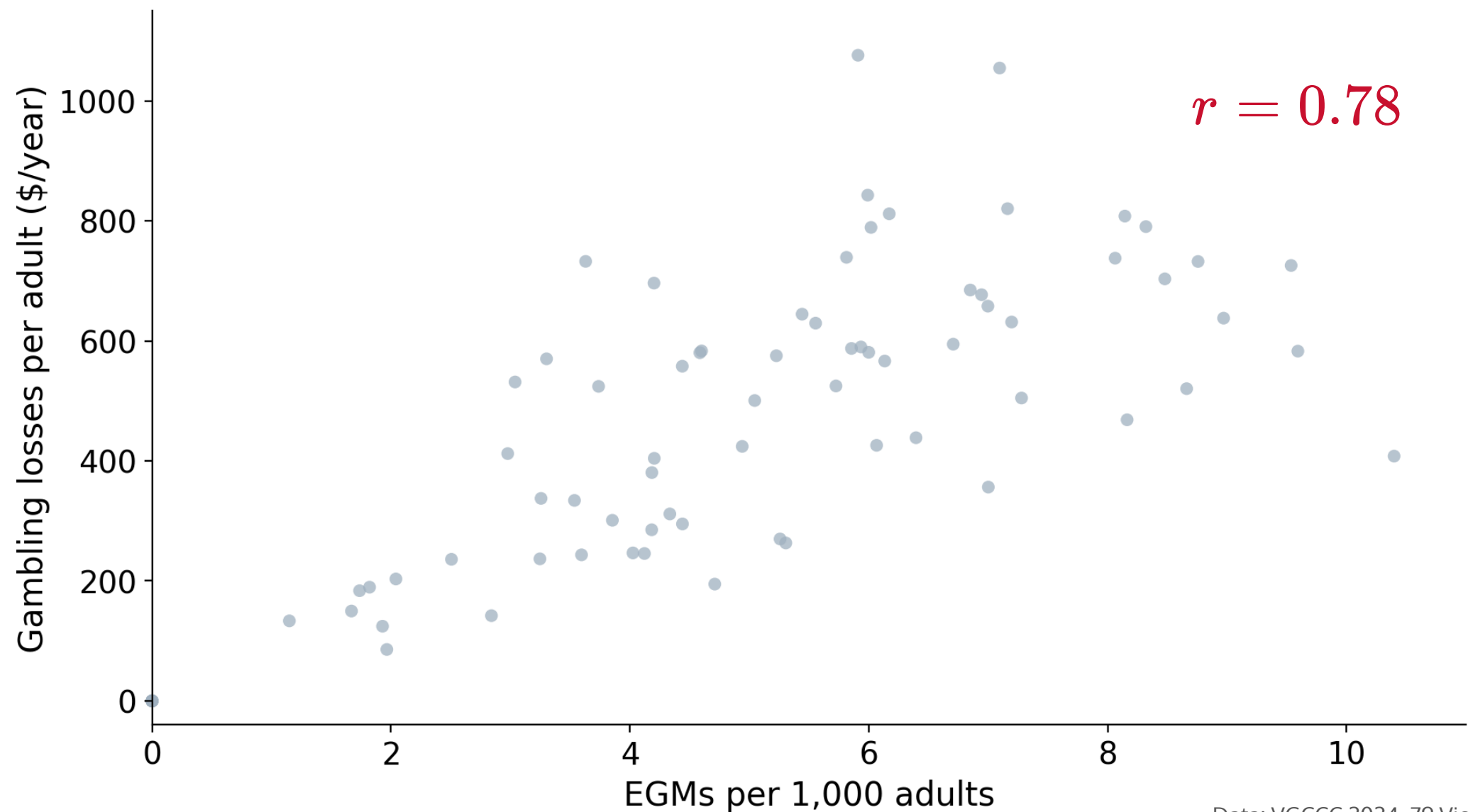
# Answer these three questions

1. How strong is the pokies-losses correlation?
2. Are losses higher where unemployment is higher?
3. Which of the three tracks losses most closely?

<https://emiliatjernstrom.com/econ2041/portal.html>

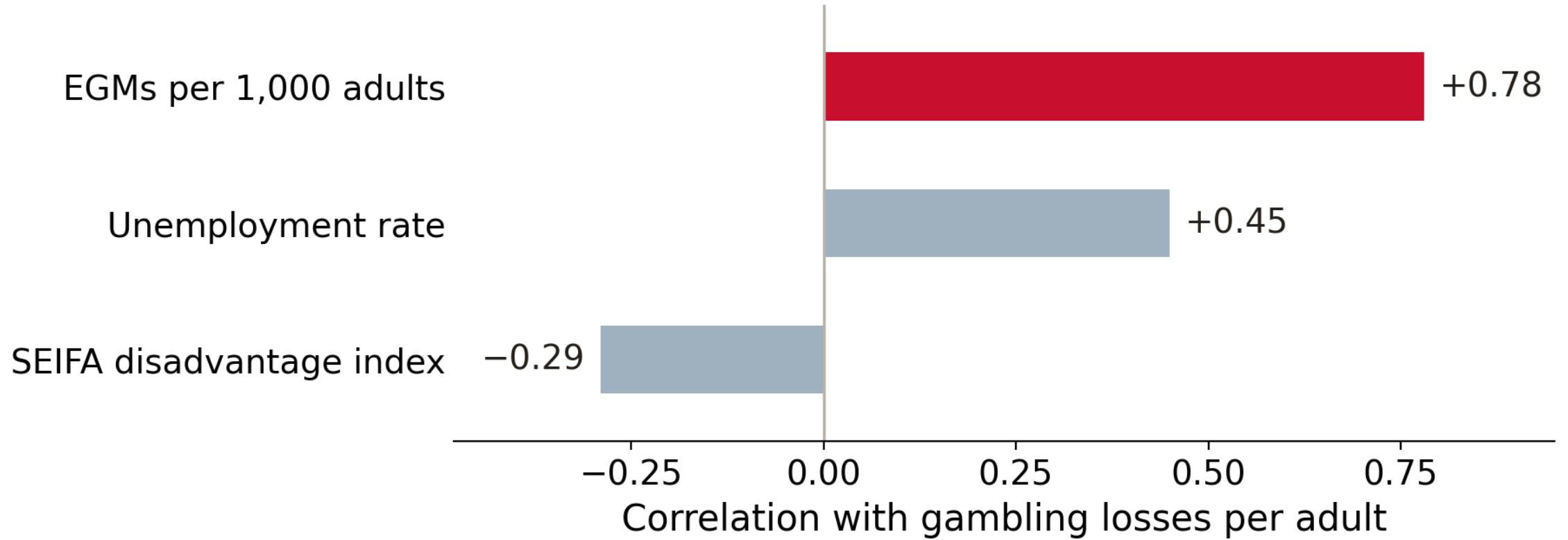


# Gambling losses vs per-adult pokies



Data: VGCCC 2024, 79 Victorian LGAs

# Pokies track losses most closely



What the correlation  
can't tell us

# You can follow along in a notebook

<https://emiliatjernstrom.com/econ2041/lectures/w06>

As usual, it will open directly in Colab

Click **Copy to Drive** first, so you can save your edits

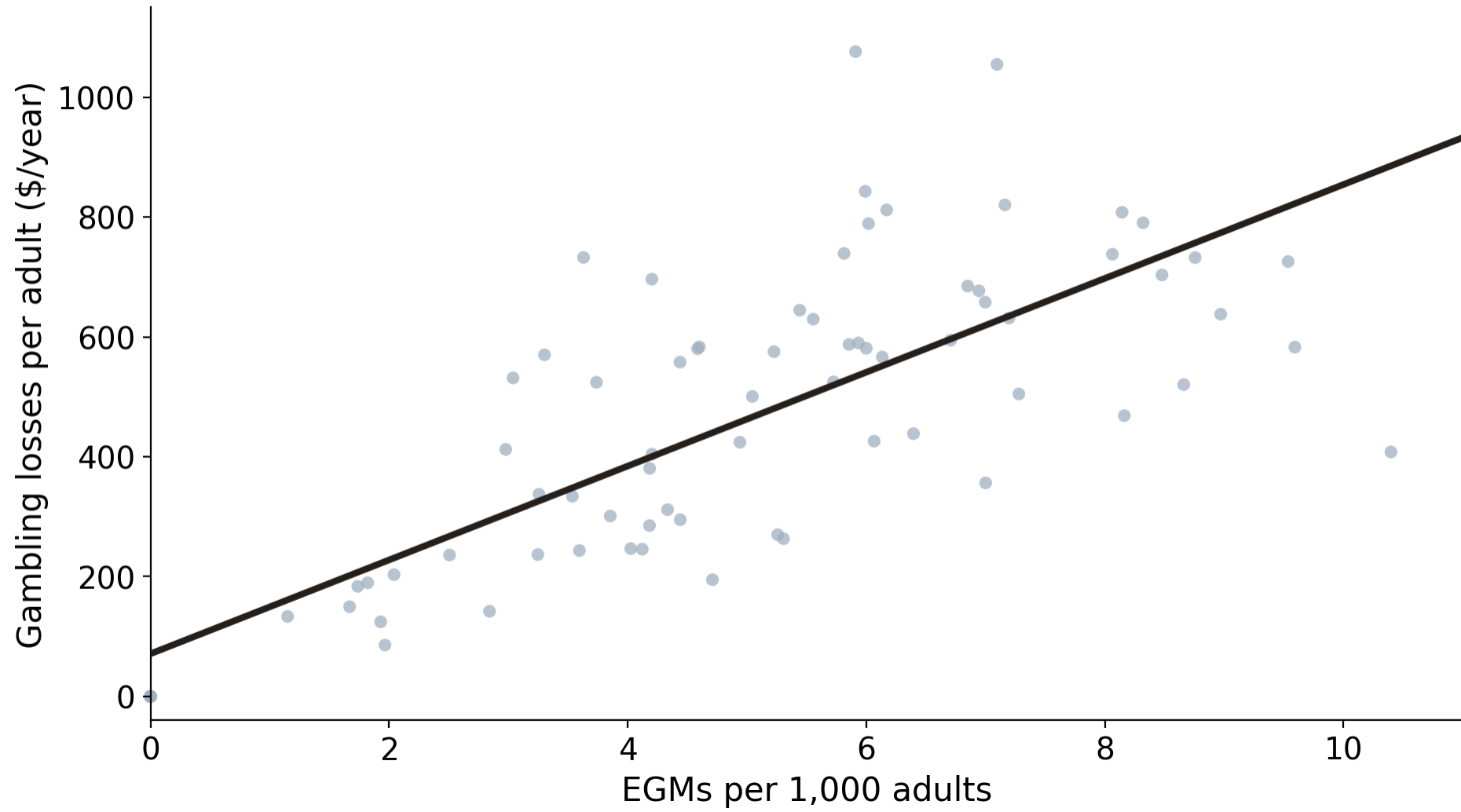
# OLS in Python

```
from statsmodels.formula.api import ols
```

```
model = ols("exp_per_adult ~ egm_per_1000", data=pokies).fit()
```

|              | coef    | std err | t      | P> t  | [0.025 | 0.975]  |
|--------------|---------|---------|--------|-------|--------|---------|
| Intercept    | 71.1722 | 38.992  | 1.825  | 0.072 | -6.471 | 148.816 |
| egm_per_1000 | 78.2968 | 7.205   | 10.866 | 0.000 | 63.949 | 92.645  |

# OLS confirms: more pokies, more losses



$$\widehat{\text{losses}} = 71 + 78 \times \text{EGMs per 1,000}$$

Data: VGCCC 2024, 79 Victorian LGAs

# Unemployment, holding pokie density fixed

Our simple regression conditioned on one variable

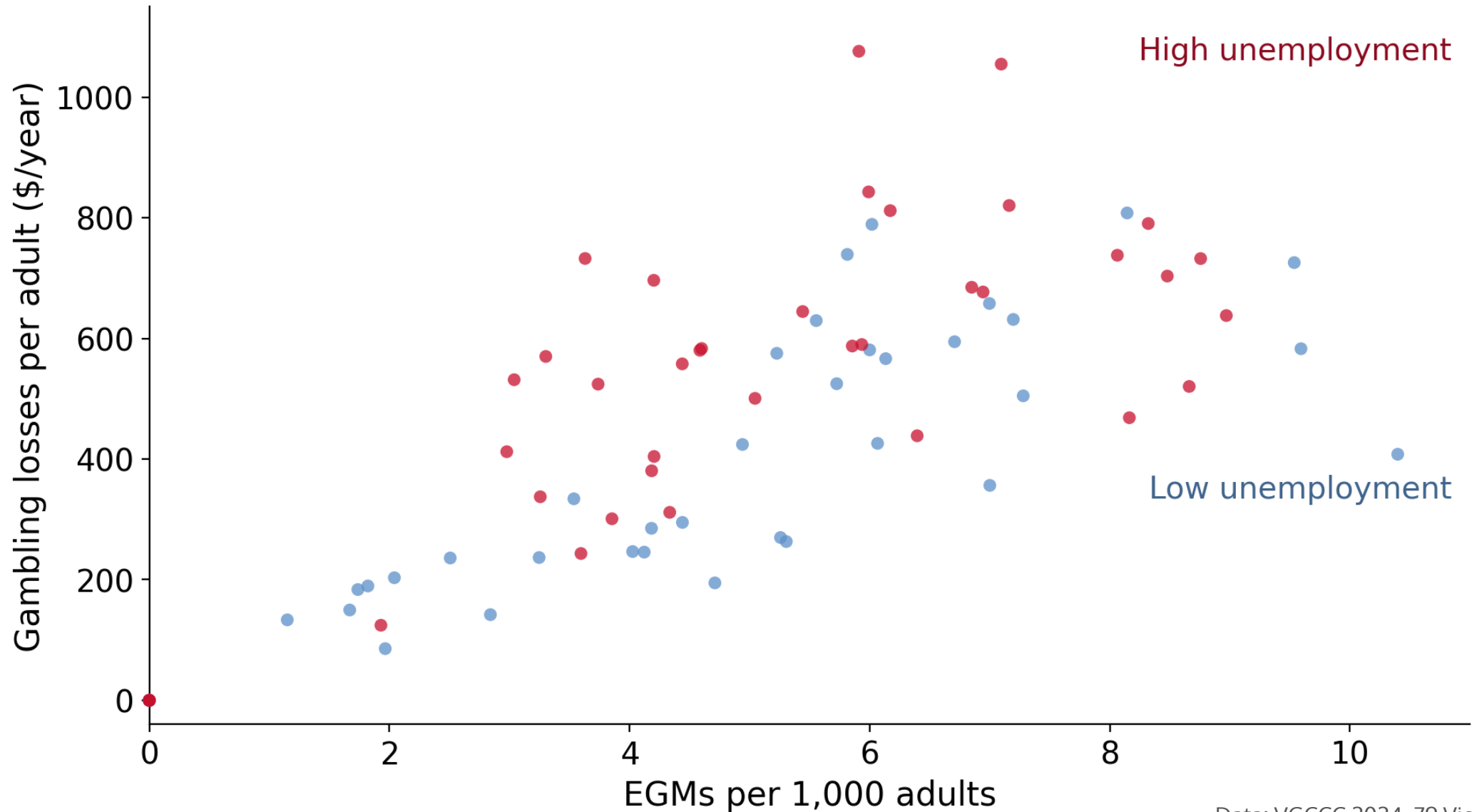
- $E[\text{losses} \mid \text{EGMs, UE} = \text{high}]$
- $E[\text{losses} \mid \text{EGMs, UE} = \text{low}]$

# Creating a categorical (dummy) variable

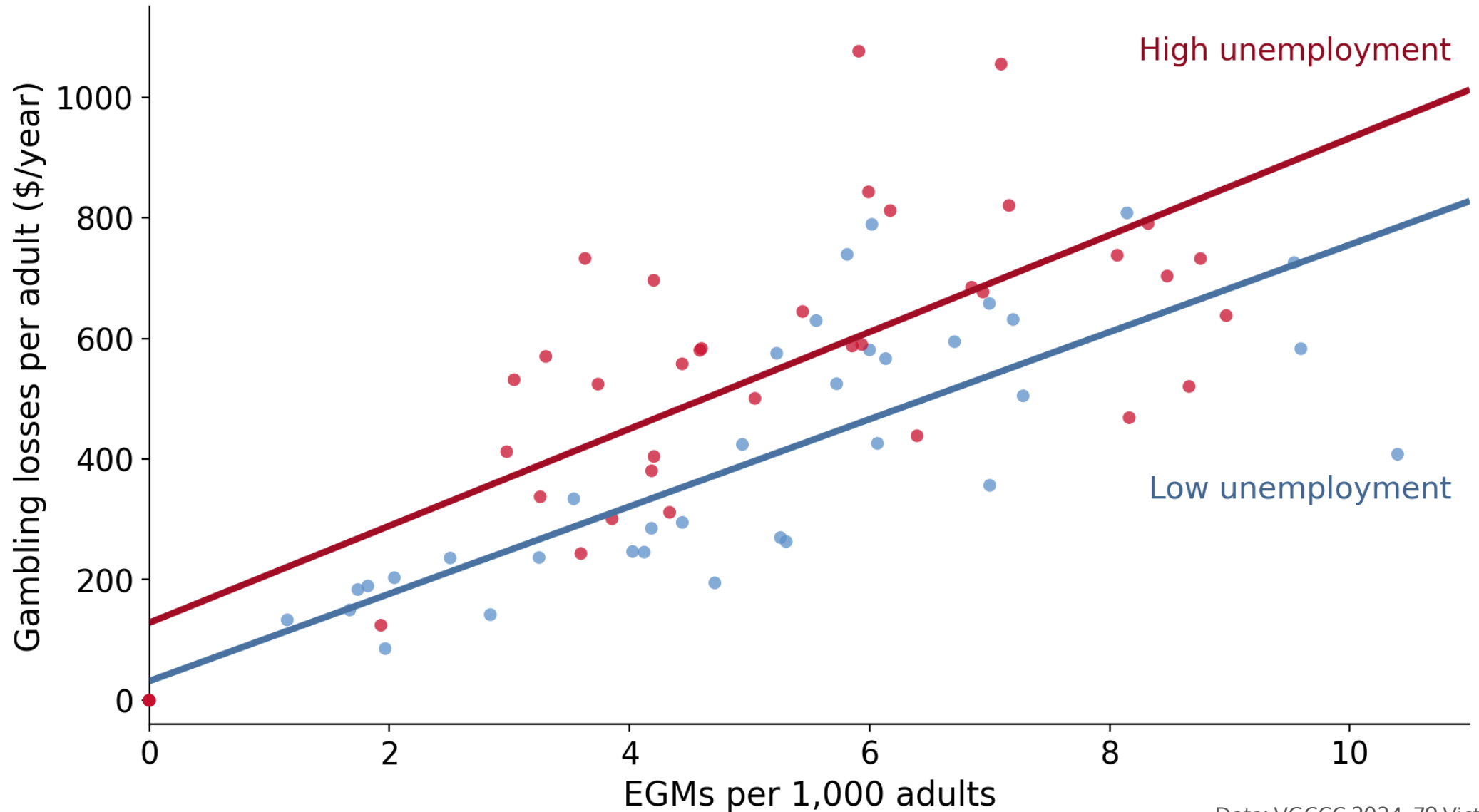
```
median_val = pokies["ue_rate"].median()  
pokies["high_ue"] = (pokies["ue_rate"] > median_val).astype(int)
```

- `high_ue` stores the answer to our comparison
- `.astype(int)` turns `True/False` into 1 and 0
- 1 = above-median unemployment (39 LGAs)
- 0 = below-median unemployment (40 LGAs)
- A 0/1 variable is called a **dummy variable**

# The same 79 LGAs, by UE category



# The same 79 LGAs, by UE category



Data: VGCCC 2024, 79 Victorian LGAs

# Multiple regression!

Both explanatory variables

One regression

All 79 LGAs

# The dummy shifts the intercept

$$\text{losses}_i = \beta_0 + \beta_1 \text{EGMs}_i + \beta_2 \text{high\_ue}_i + u_i$$

For low-unemployment LGAs ( $\text{high\_ue} = 0$ ): intercept  $\beta_0$

For high-unemployment LGAs ( $\text{high\_ue} = 1$ ): intercept  $\beta_0 + \beta_2$

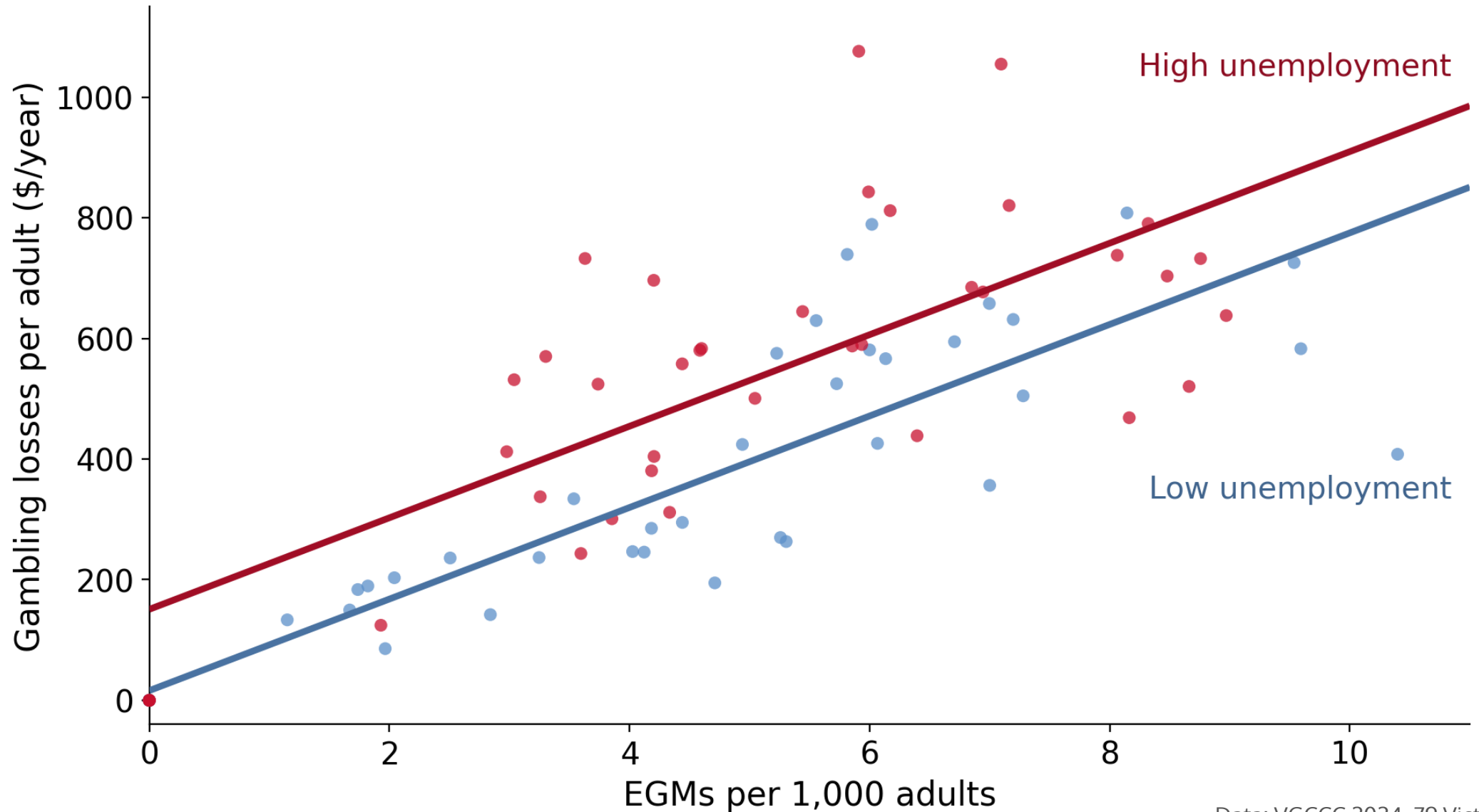
Same slope  $\beta_1$  in both groups: two parallel lines

# Add `high_ue` to the ols regression!

```
ols("exp_per_adult ~ egm_per_1000 + high_ue", data=pokies).fit()
```

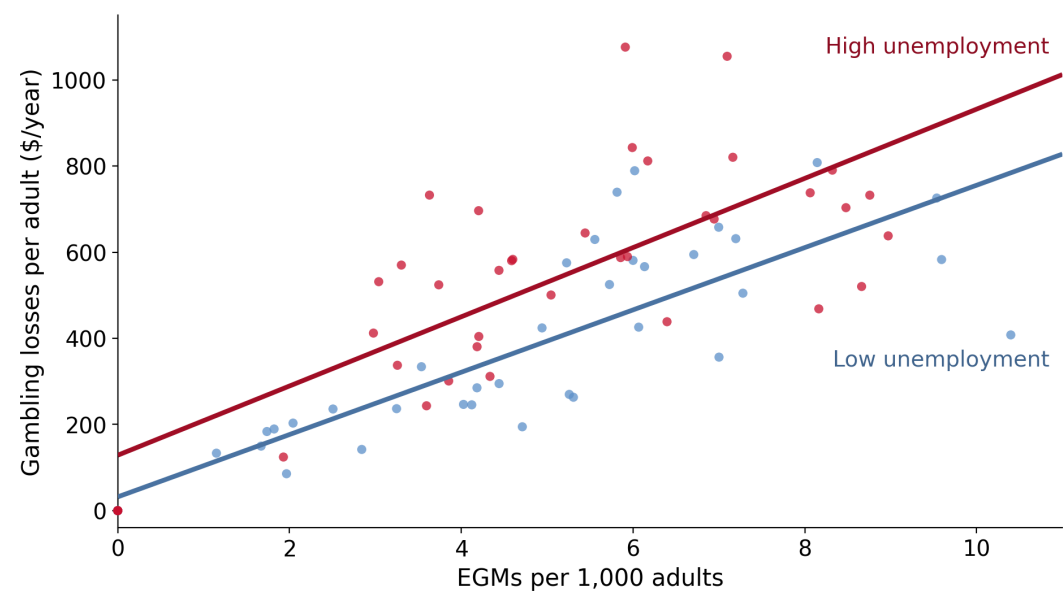
|              | coef     | std err | t      | P> t  | [0.025  | 0.975]  |
|--------------|----------|---------|--------|-------|---------|---------|
| Intercept    | 16.1362  | 38.739  | 0.417  | 0.678 | -61.018 | 93.291  |
| egm_per_1000 | 75.8681  | 6.674   | 11.367 | 0.000 | 62.575  | 89.161  |
| high_ue      | 134.7112 | 35.300  | 3.816  | 0.000 | 64.406  | 205.016 |

# Two parallel lines, shifted

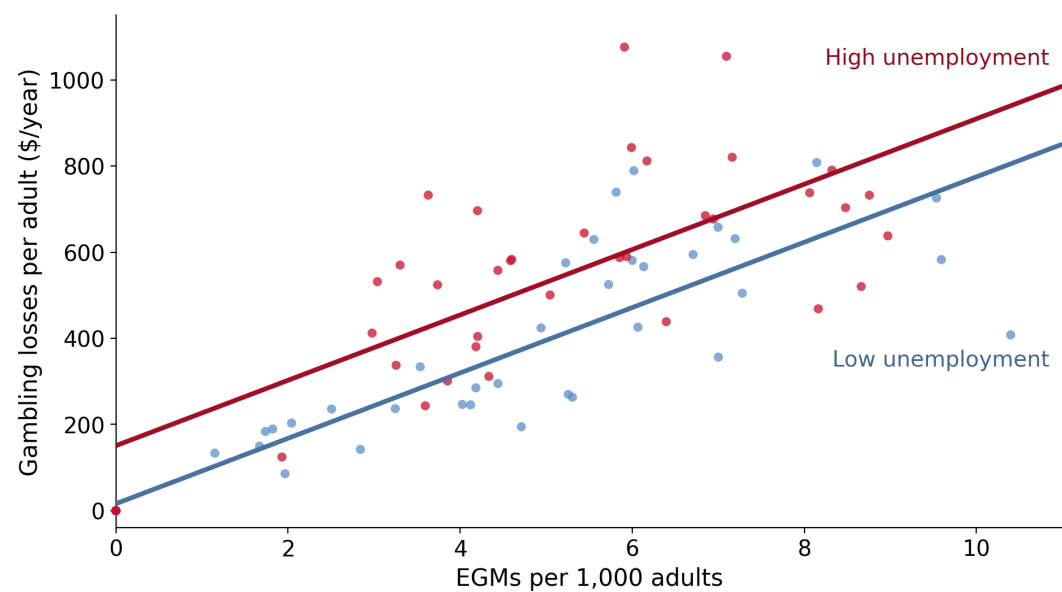


# Split sample vs shared slope

Two regressions, two slopes



One regression, one shared slope



# The second X can be continuous

```
pokies["ue_pct"] = 100 * pokies["ue_rate"]  
ols("exp_per_adult ~ egm_per_1000 + ue_pct", data=pokies).fit()
```

|              | coef      | std err | t      | P> t  | [0.025   | 0.975]  |
|--------------|-----------|---------|--------|-------|----------|---------|
| Intercept    | -115.9316 | 49.520  | -2.341 | 0.022 | -214.559 | -17.304 |
| egm_per_1000 | 72.8291   | 6.329   | 11.507 | 0.000 | 60.224   | 85.434  |
| ue_pct       | 61.1848   | 11.845  | 5.165  | 0.000 | 37.593   | 84.776  |

# What multivariate regression buys us

Uses **all** the data for every coefficient → more precise estimates

Holds the other X fixed: but only the variables we put in the formula

Scales: if we have 3, 4, or 10 X-variables, splitting would not work!

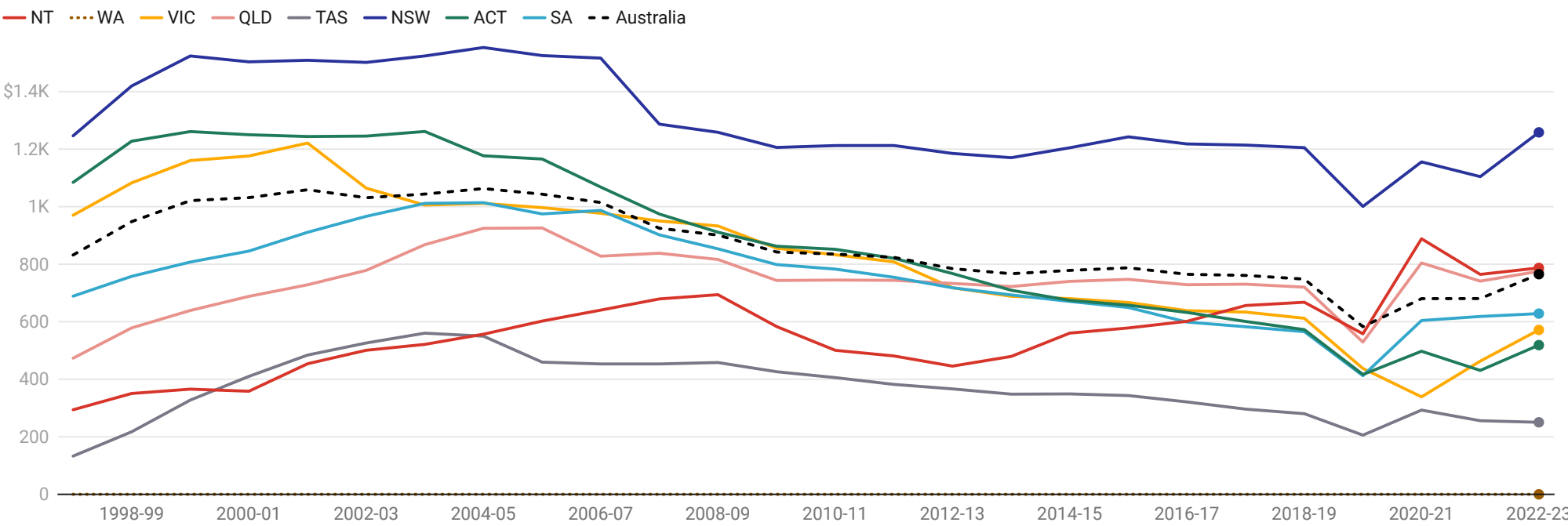
# Some take-aways

- A categorical variable enters regression as a 0/1 "dummy variable"
- The dummy shifts our regression line
- Multivariate regression holds the other explanatory variables fixed & uses all the data

# Losses track where pokies are allowed

## Average per capita losses on gaming machines across Australia

Gaming machines include slots, pokies, poker machines, and fruit machines



Based on real per capita expenditure, which adjusts for inflation and population growth  
Chart: The Conversation • Source: [Queensland Government Statistician's Office](#) • [Get the data](#) • [Embed](#) • [Download image](#) • Created with [Datawrapper](#)