FY22 Budget Training

using and generating evidence

David_Yokum@brown.edu

Director, The Policy Lab
thepolicylab.brown.edu

August 27, 2020
Rhode Island

learning objectives

- How do you assess the existing evidence base?
  - Methods 101 crash course
  - The Evidence Scale
- When and how might you invest in generating new evidence?

- Recall other resources:
  - Budget Instructions
  - Decision Package Template
  - Examples (good and bad)
  - Website Trainer
  - Office Hours

what evidence do you need?

3. Opportunity Statement

In this section, clearly explain the problem that exists today and the opportunity that your request presents to capitalize on. The best opportunity statements thoroughly explain, with as much detail as possible: (1) where we are today; (2) where we want to be in the future; and (3) why there is the gap between where we are and where we want to be. The best opportunity statements also quantify key variables wherever possible.

4. Proposed Intervention & Theory of Change

Provide a detailed description of the initiative you are proposing to respond to the above-described problem/capitalize on the opportunity. Your narrative here should clearly describe how your intervention, if funded, could close the gap described above and achieve the desired future state.

types of metrics

inputs	outputs	outcomes	impacts

types of metrics

inputs	**outputs**	**outcomes**	**impacts**
staff facilities materials etc.	# people served # job apps submitted	# employed average wages	Increase in employment or wages CAUSED BY the program

aims of evaluation

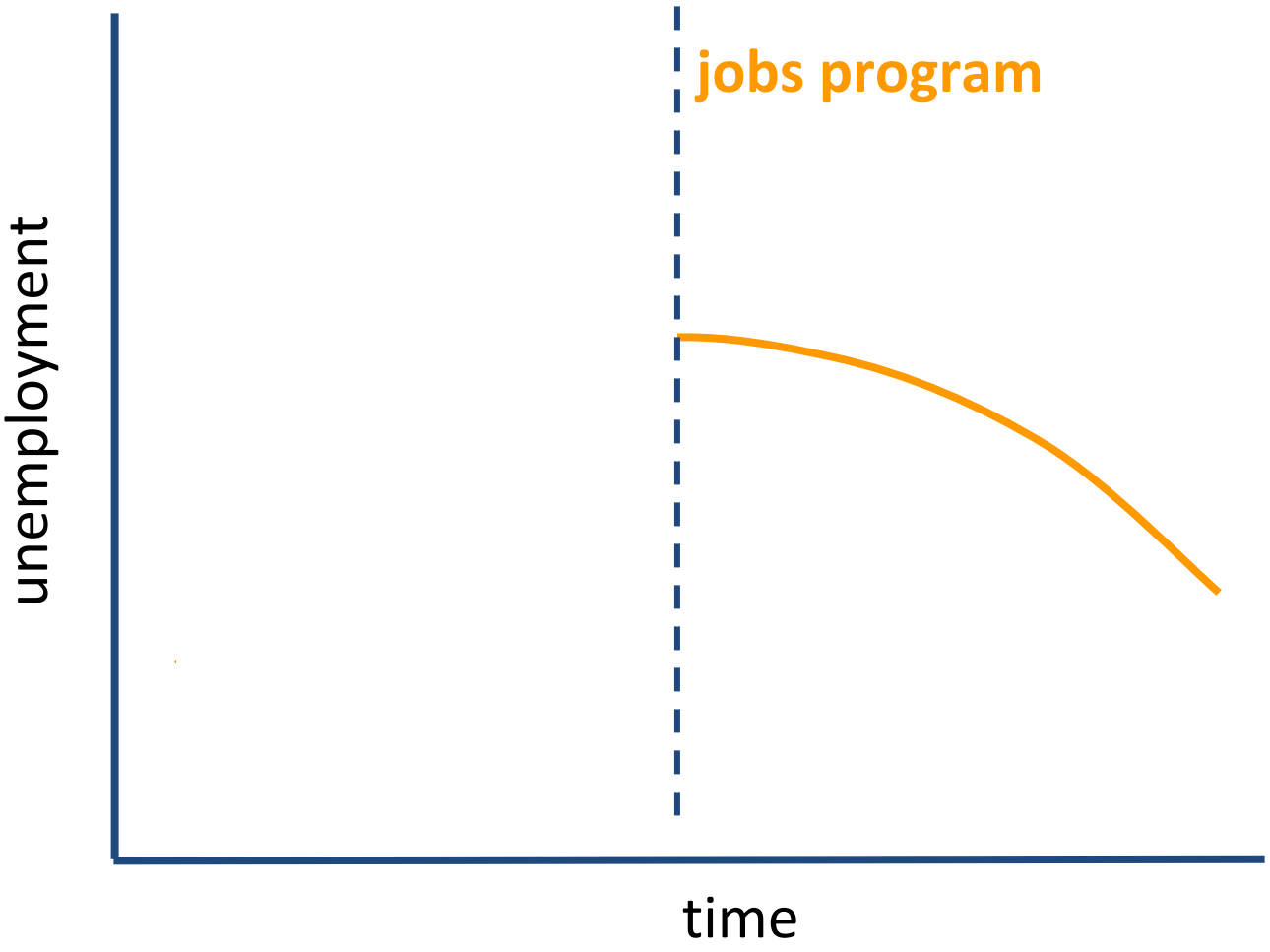
inputs	outputs	outcomes	impacts
staff facilities materials etc.	# people served # job apps submitted	# employed average wages	Increase in employment or wages CAUSED BY the program

implementation

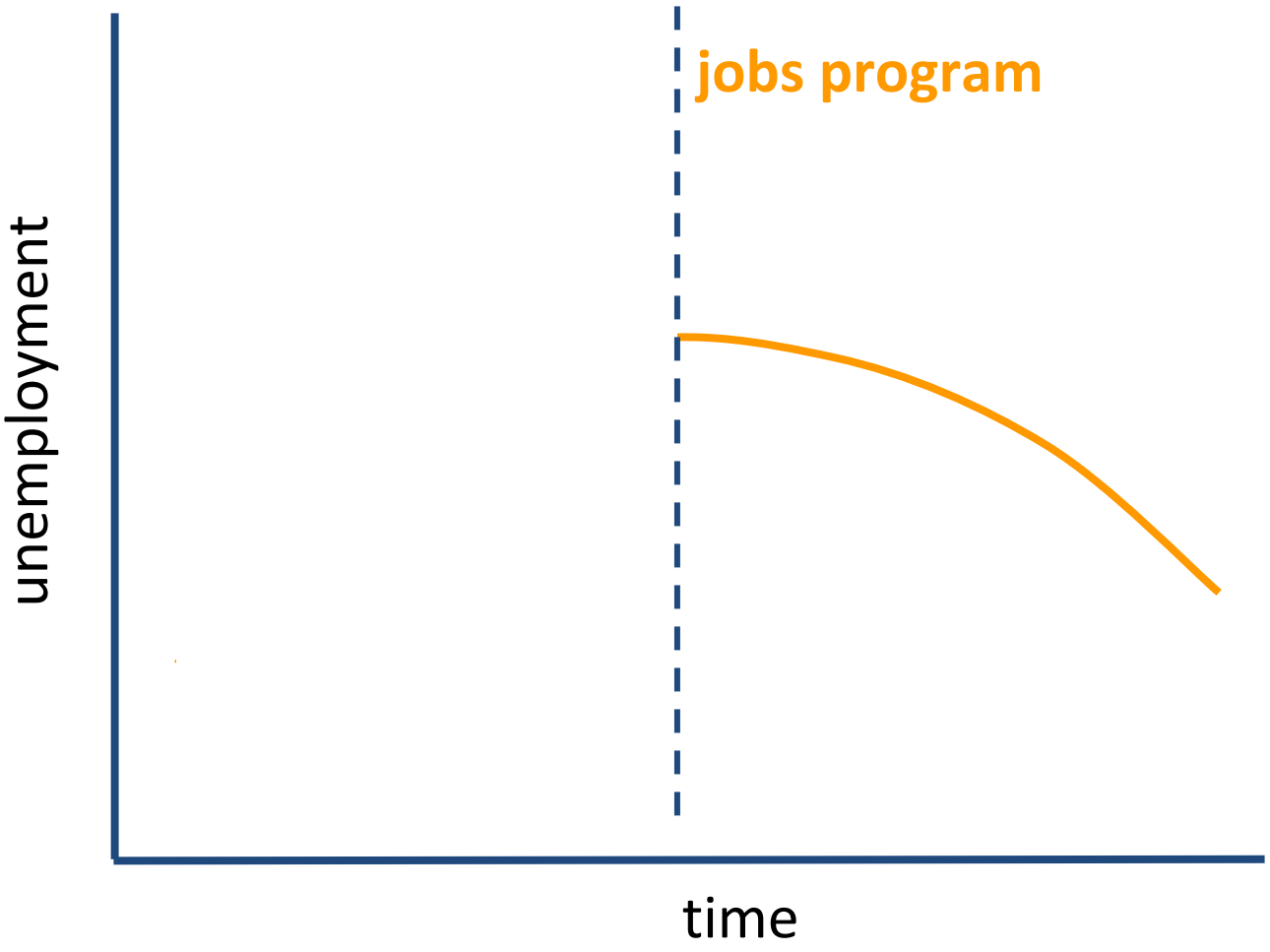
outcome

impact

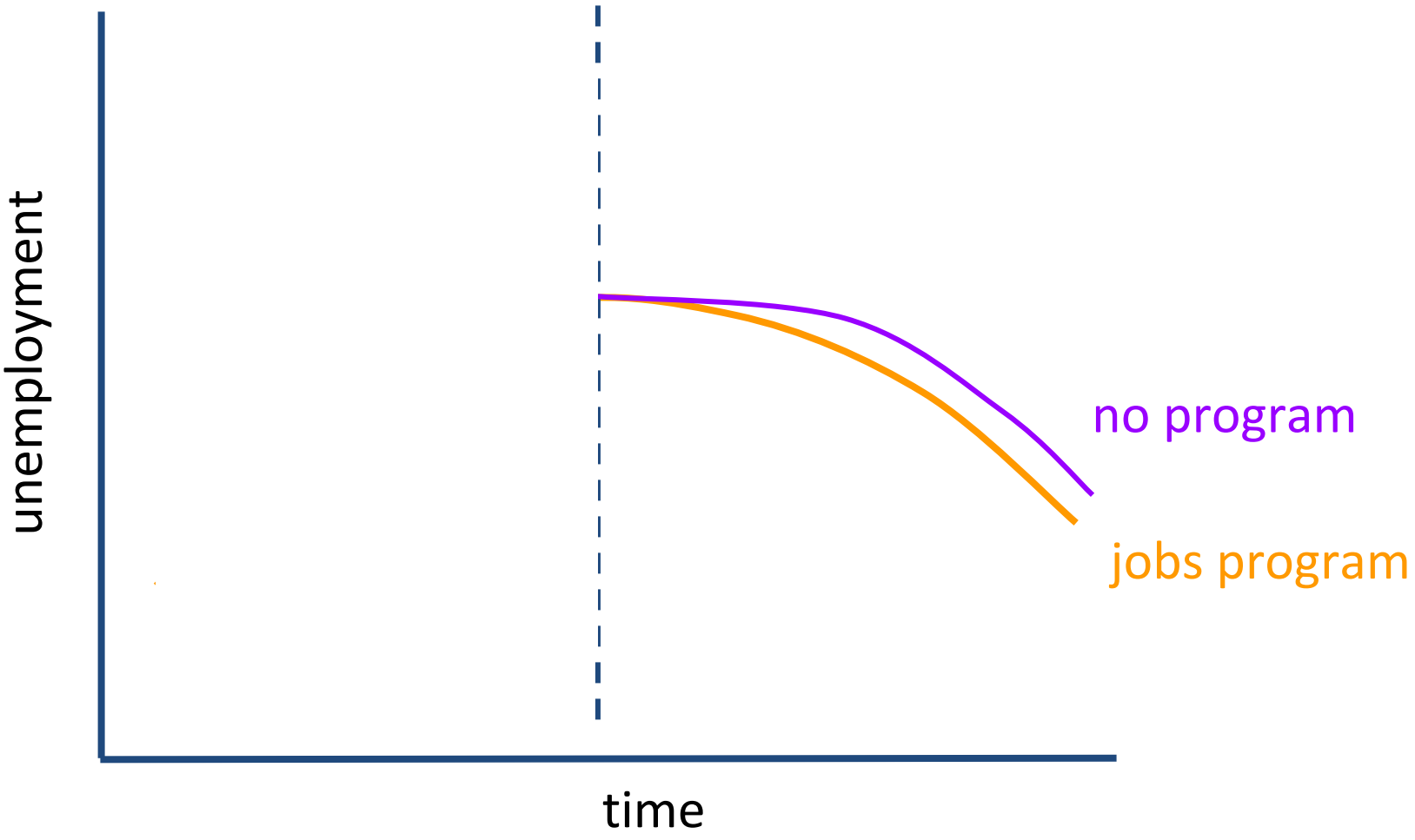
impact / causal claims



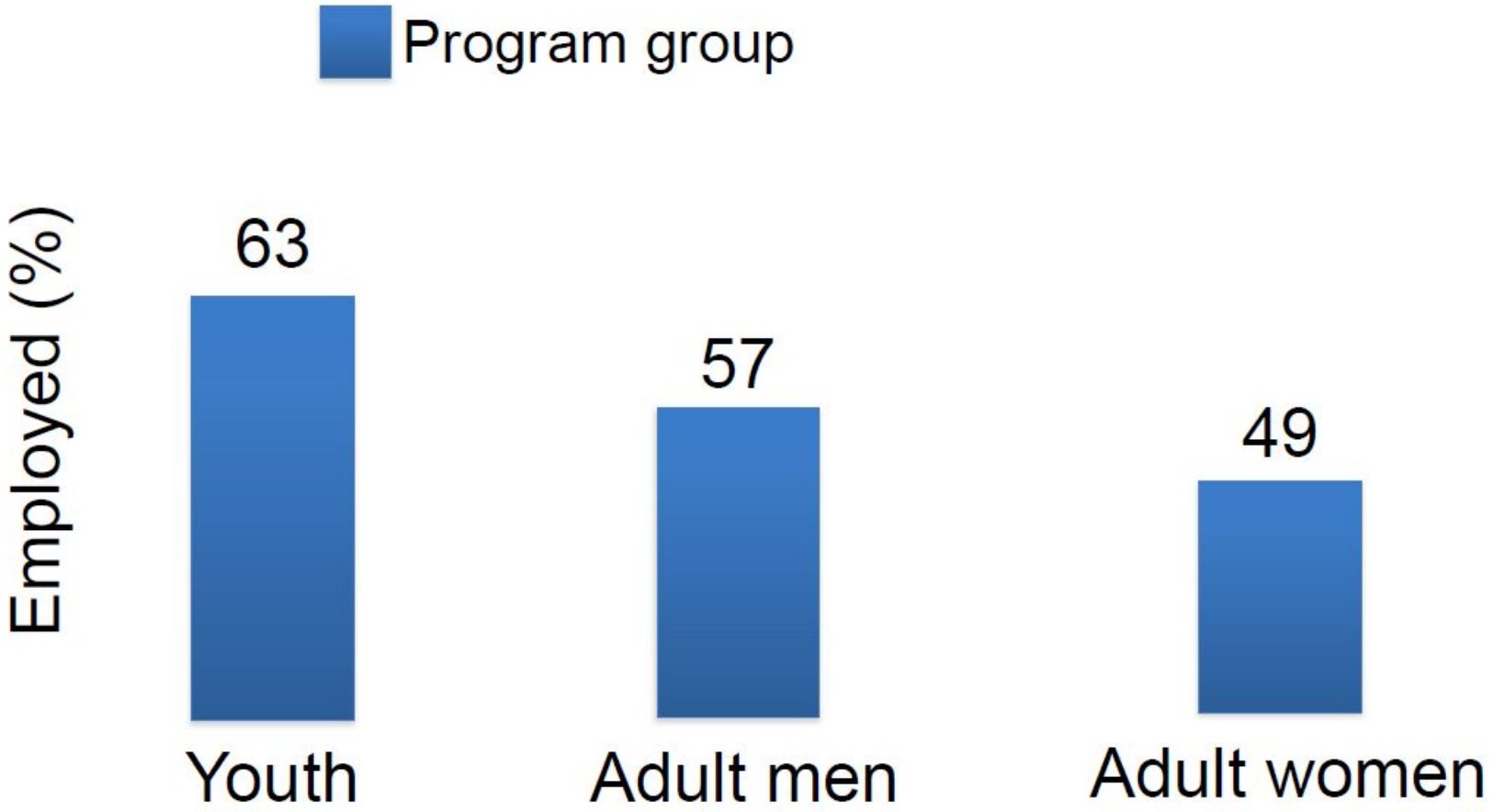
CAUSAL CLAIMS



need a counterfactual

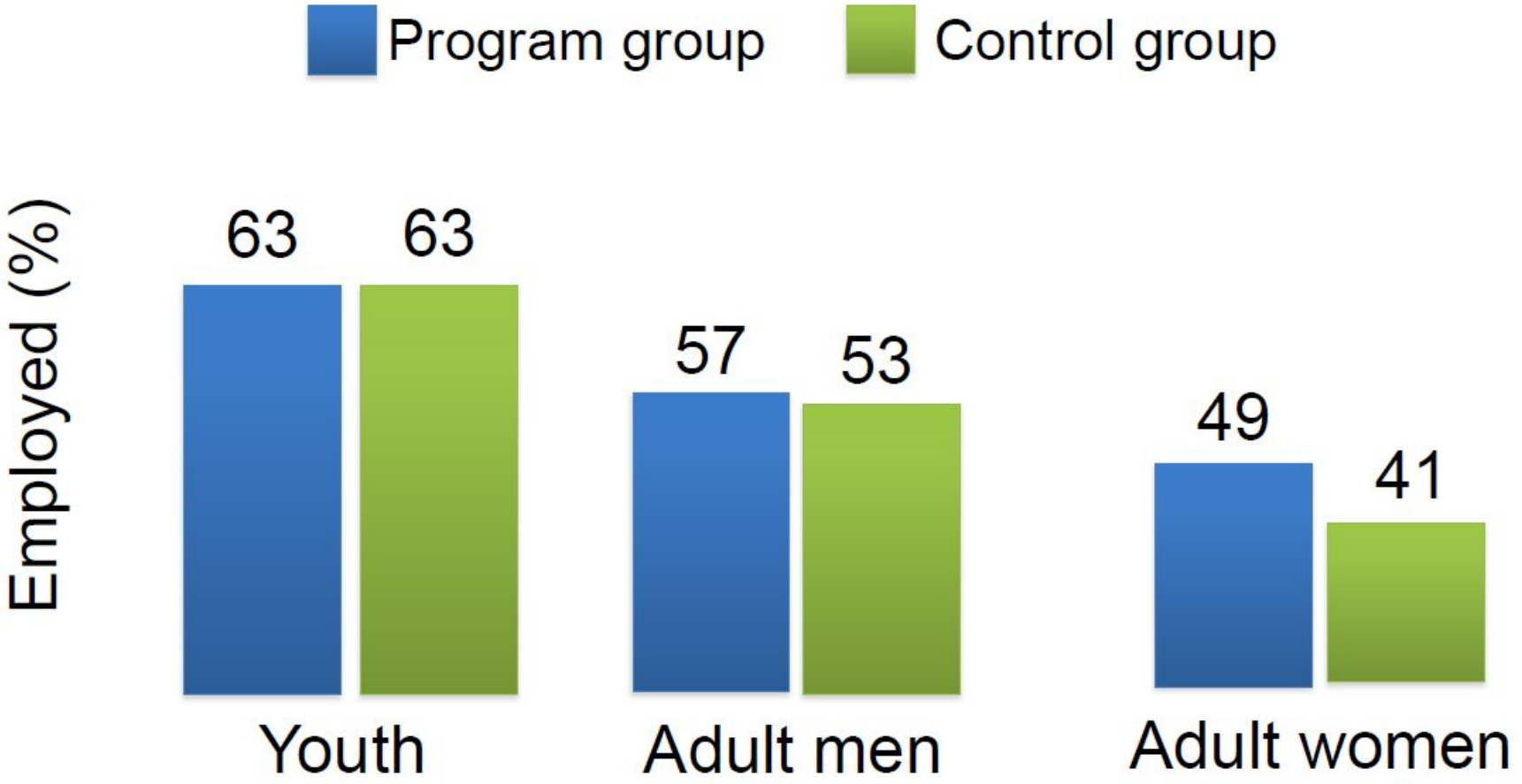


who benefited the most from the jobs program?



Aack, I can't find the publication citation for this data!

who benefited the most from the jobs program?

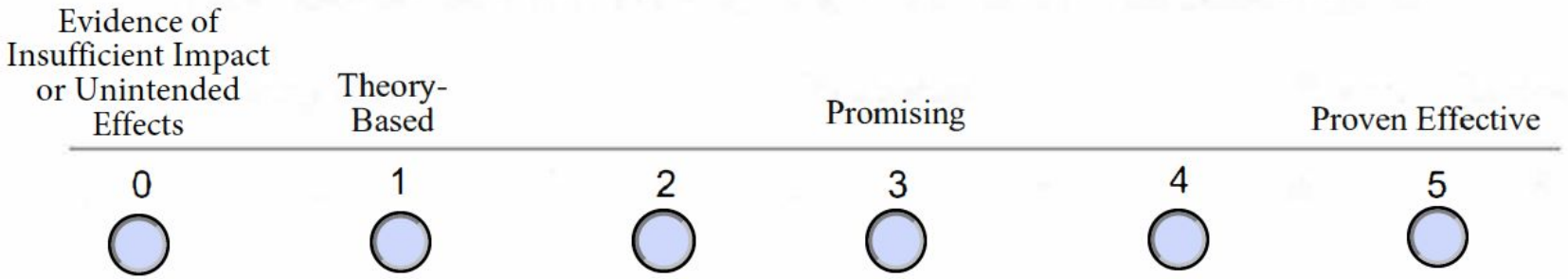


Aack, I can't find the publication citation for this data!

finding the evidence you need

6. Evidence Scale Ranking

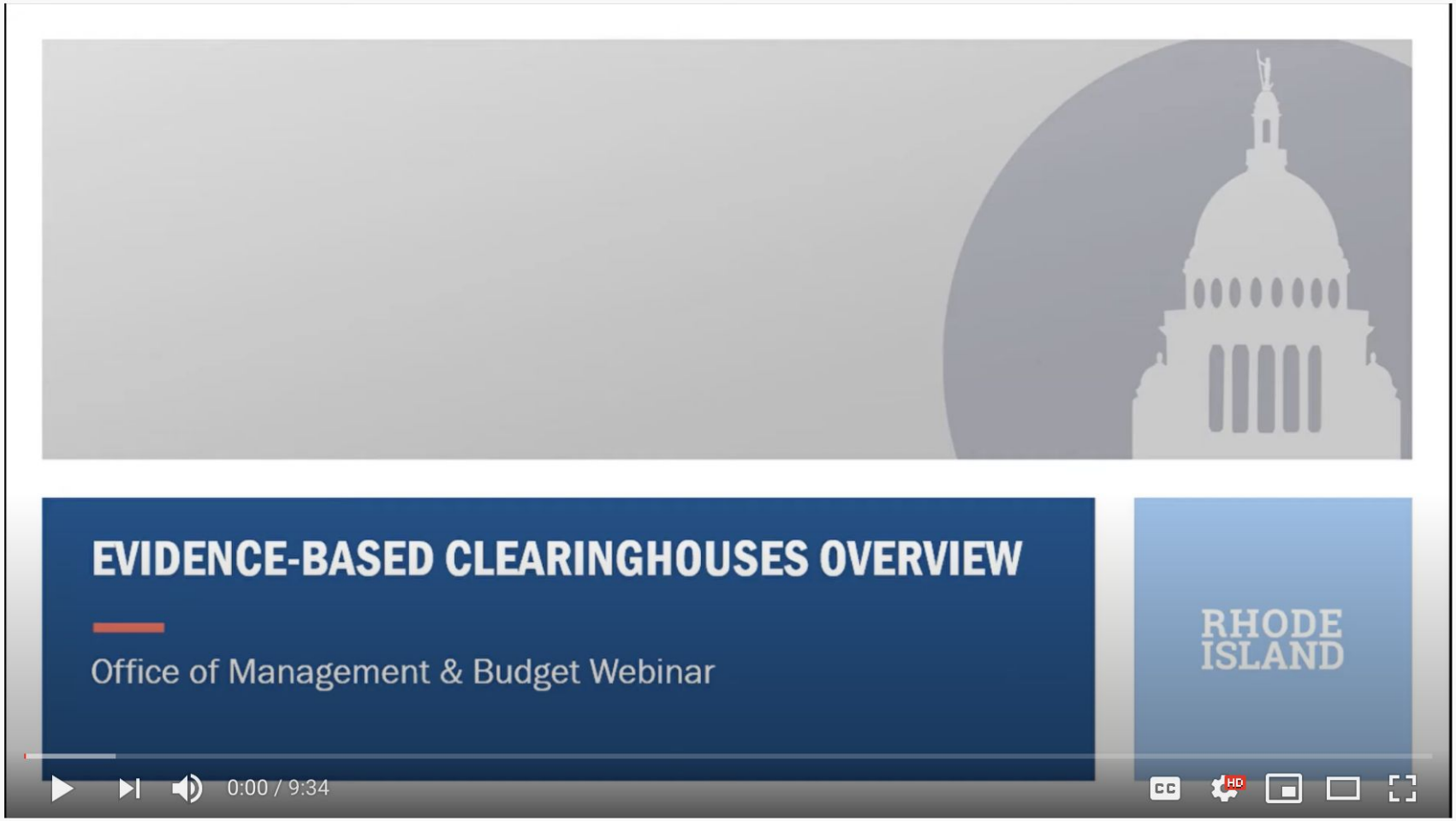
Please rank the proposed initiative's current level of evidentiary support on a scale from 0-5, based on the RI Evidence Scale, with one being the least evidentiary support and five being the most evidentiary support. You can use tools like the Pew Results First Clearinghouse and the Social Programs That Work database to determine whether the initiative you are proposing has been rigorously evaluated in other jurisdictions. The Office of Management & Budget understands that the majority of agency requests will likely not be in the top evidence tiers at the point of submittal, and you should certainly feel free to submit requests that are "theory-based" rather than evidence based. Please note that "theory-based" submissions should include a robust and compelling measurement and evaluation plan in the Performance Measurement section.



7. Description of Evidence Base

Describe the justification for your evidence scale ranking. What evidence exists that makes you think that the proposed initiative will work? Where is there uncertainty of effectiveness? It is helpful to include citations, links, or attachments of relevant evidence source(s)

evidence clearinghouses



https://www.youtube.com/watch?v=AFuOslr2cSw&feature=youtu.be

11 Tips for Doing Desk Research

WARNING: study quality

- Too few people were studied
- The people studied are not representative of the population you care about
- Causal claims made, but no counterfactual
- Effect sizes are not talked about clearly
- Fishy handling of the data

WARNING: study quality

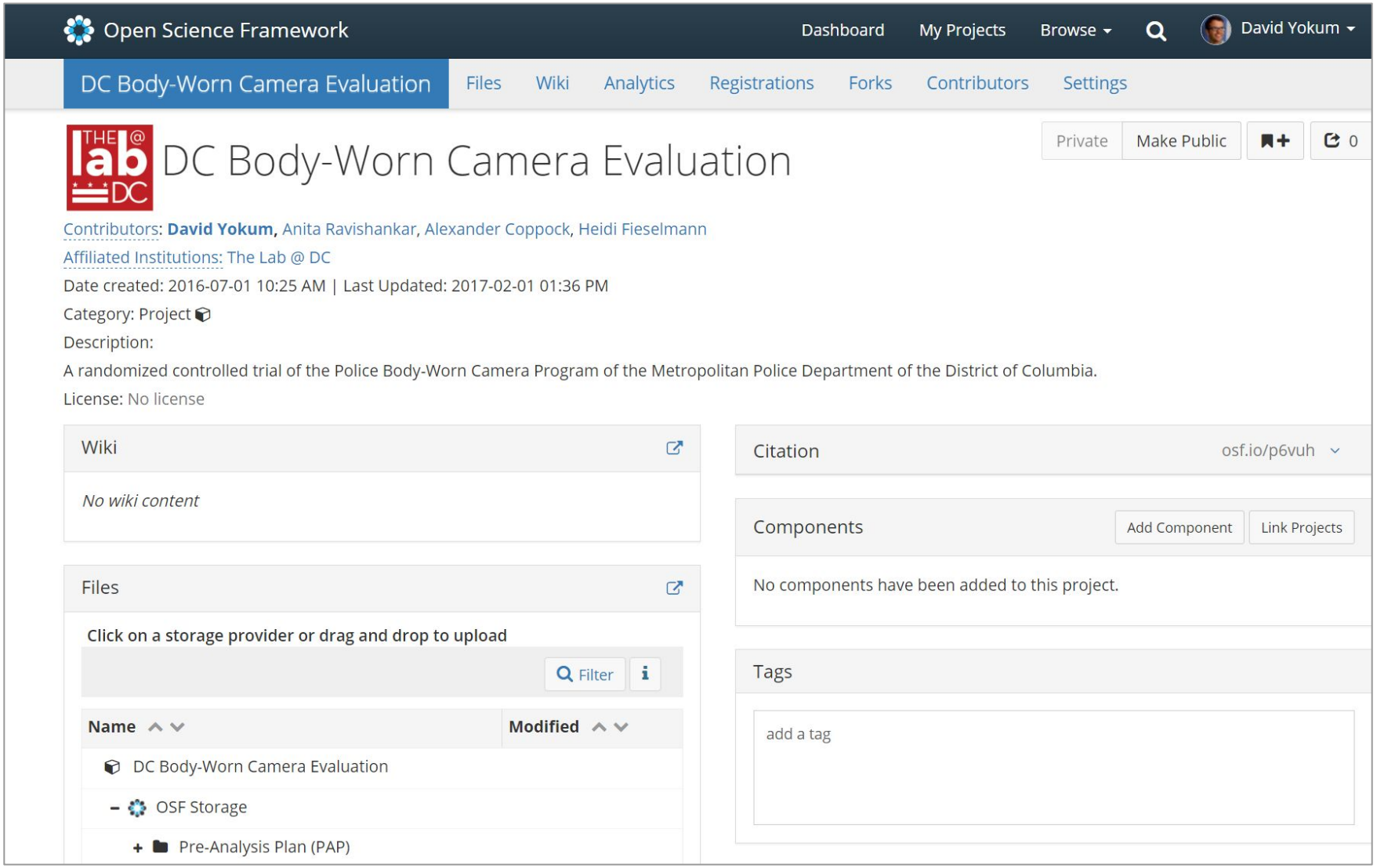
- Too few people were studied
- The people studied are not representative of the population you care about
- Causal claims made, but no counterfactual
- Effect sizes are not talked about clearly
- **Fishy handling of the data**

fishy handling of data

- File drawer problem
- *P*-hacking

See generally, Lindsay, Simons, & Lilienfeld, "Research Preregistration 101," APS Observer (December 2016).

pre-analysis plans help avoid fishiness



See https://osf.io/yjyng/ for the publically pre-registered analysis plan. See also https://osf.io/q6c45/ for PVD Talks example.

methods crash course

9. What methodologies do you currently use for program evaluation? Check all that apply.

- ☐ No evaluation is done
- ☐ Measurement of the resources (e.g. staff, material expenses) required to deploy the service or initiative
- ☐ Measurement of how many people use the service or initiative
- ☐ Focus groups, surveys, or other qualitative methods that ask people about their experience with the service/initiative
- ☐ Measurement of outcomes at a point in time or over time (e.g. monthly or quarterly reports of student test scores, average wages, crime reports, park admissions, medical claims, etc.)
- ☐ Measurement of causal impact with econometric methods (e.g. regression discontinuity, multiple regression, matched controls, instrumental variables)
- ☐ Measurement of causal impact with a randomized control trial (RCT)
- ☐ Other:

inputs

9. What methodologies do you currently use for program evaluation? Check all that apply.

- ☐ No evaluation is done
- ☐ Measurement of the resources (e.g. staff, material expenses) required to deploy the service or initiative
- ☐ Measurement of how many people use the service or initiative
- ☐ Focus groups, surveys, or other qualitative methods that ask people about their experience with the service/initiative
- ☐ Measurement of outcomes at a point in time or over time (e.g. monthly or quarterly reports of student test scores, average wages, crime reports, park admissions, medical claims, etc.)
- ☐ Measurement of causal impact with econometric methods (e.g. regression discontinuity, multiple regression, matched controls, instrumental variables)
- ☐ Measurement of causal impact with a randomized control trial (RCT)
- ☐ Other:

outputs

9. What methodologies do you currently use for program evaluation? Check all that apply.

- ☐ No evaluation is done
- ☐ Measurement of the resources (e.g. staff, material expenses) required to deploy the service or initiative
- ☐ Measurement of how many people use the service or initiative
- ☐ Focus groups, surveys, or other qualitative methods that ask people about their experience with the service/initiative
- ☐ Measurement of outcomes at a point in time or over time (e.g. monthly or quarterly reports of student test scores, average wages, crime reports, park admissions, medical claims, etc.)
- ☐ Measurement of causal impact with econometric methods (e.g. regression discontinuity, multiple regression, matched controls, instrumental variables)
- ☐ Measurement of causal impact with a randomized control trial (RCT)
- ☐ Other:

outcomes

9. What methodologies do you currently use for program evaluation? Check all that apply.

- ☐ No evaluation is done
- ☐ Measurement of the resources (e.g. staff, material expenses) required to deploy the service or initiative
- ☐ Measurement of how many people use the service or initiative
- ☐ Focus groups, surveys, or other qualitative methods that ask people about their experience with the service/initiative
- ☐ Measurement of outcomes at a point in time or over time (e.g. monthly or quarterly reports of student test scores, average wages, crime reports, park admissions, medical claims, etc.)
- ☐ Measurement of causal impact with econometric methods (e.g. regression discontinuity, multiple regression, matched controls, instrumental variables)
- ☐ Measurement of causal impact with a randomized control trial (RCT)
- ☐ Other:

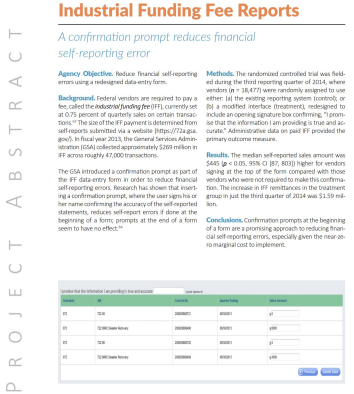
impacts

9. What methodologies do you currently use for program evaluation? Check all that apply.

- ☐ No evaluation is done
- ☐ Measurement of the resources (e.g. staff, material expenses) required to deploy the service or initiative
- ☐ Measurement of how many people use the service or initiative
- ☐ Focus groups, surveys, or other qualitative methods that ask people about their experience with the service/initiative
- ☐ Measurement of outcomes at a point in time or over time (e.g. monthly or quarterly reports of student test scores, average wages, crime reports, park admissions, medical claims, etc.)
- ☐ Measurement of causal impact with econometric methods (e.g. regression discontinuity, multiple regression, matched controls, instrumental variables)
- ☐ Measurement of causal impact with a randomized control trial (RCT)
- ☐ Other:

wait, can't folk just self-report impact?

1. People may answer in ways they think *you* want, rather than what *they* really think.
2. People can't report unconscious causes of behavior.
3. Memory is faulty.
4. Often not a sufficiently representative sample.



PROJECT ABSTRACT

Industrial Funding Fee Reports

A confirmation prompt reduces financial self-reporting error

Agency Objective. Reduce financial self-reporting errors using a redesigned data-entry form.

Background. Federal vendors are required to pay a fee, called the *industrial funding fee* (IFF), currently set at 0.75 percent of quarterly sales on certain transactions.[1] The size of the IFF payment is determined from self-reports submitted via a website (https://73a.gsa.gov). In fiscal year 2013, the General Services Administration (GSA) collected approximately $269 million in IFF across roughly 47,000 transactions.

The GSA introduced a confirmation prompt as part of the IFF data-entry form in order to reduce financial self-reporting errors. Research has shown that inserting a confirmation prompt, where the user signs his or her name confirming the accuracy of the self-reported statements, reduces self-report errors if done at the beginning of a form; prompts at the end of a form seem to have no effect.[2]

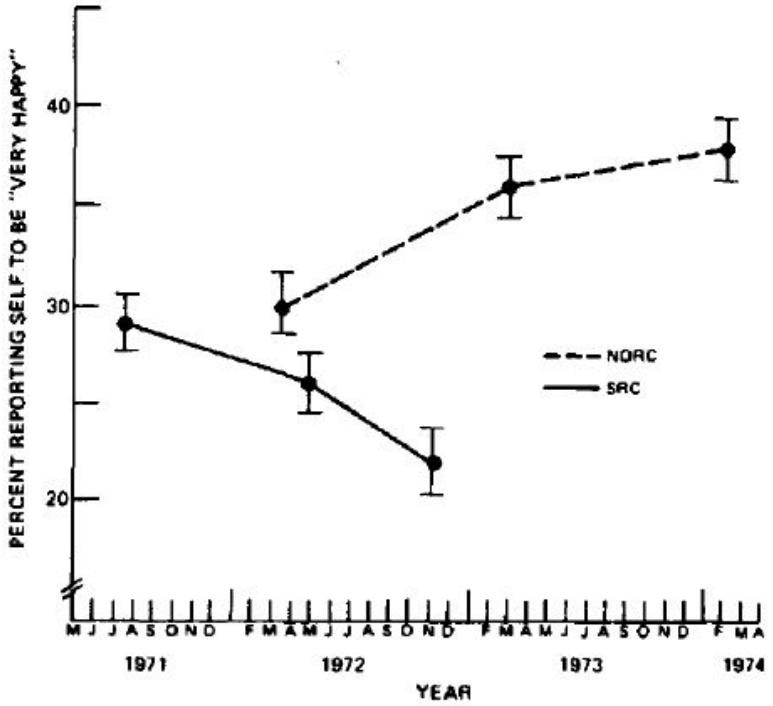
Methods. The randomized controlled trial was fielded during the third reporting quarter of 2014, where vendors (n = 18,477) were randomly assigned to use either: (a) the existing reporting system (control); or (b) a modified interface (treatment), redesigned to include an opening signature box confirming, "I promise that the information I am providing is true and accurate." Administrative data on paid IFF provided the primary outcome measure.

Results. The median self-reported sales amount was $445 ($p < 0.05$, 95% CI [87, 803]) higher for vendors signing at the top of the form compared with those vendors who were not required to make this confirmation. The increase in IFF remittances in the treatment group in just the third quarter of 2014 was $1.59 million.

Conclusions. Confirmation prompts at the beginning of a form are a promising approach to reducing financial self-reporting errors, especially given the near-zero marginal cost to implement.

See e.g. GSA Office of Evaluation Sciences, "Reducing Self-reporting Errors by IFF Form Users."

Figure 1
Trends in Self-Reported Happiness, 1971–1973



Note. Estimates are derived from sample surveys of noninstitutionalized population of the continental United States, aged 18 and over. Error bars demark ±1 standard error around sample estimate. NORC = National Opinion Research Center; SRC = Survey Research Center. Questions were "Taken all together, how would you say things are these days—would you say that you are very happy, pretty happy, or not too happy?" (NORC); and "Taking all things together, how would you say things are these days—would you say you're very happy, pretty happy, or not too happy these days?" (SRC). From "Why Do Surveys Disagree? Some Preliminary Hypotheses and Some Disagreeable Examples" (p. 166) by C. F. Turner, 1984, in C. F. Turner and E. Martin, *Surveying Subjective Phenomena*, New York: Russell Sage Foundation. Copyright 1984 by the Russell Sage Foundation. Reprinted by permission.

family context

work context

Fischhoff (1991). Value elicitation: Is there anything in there? *Ame. Psych.*, *46*(8): 835-847.

causal methodologies

1. randomized controlled trials (RCTs)
2. “natural” experiments
3. pre-post comparisons
4. multiple regression / matching
5. instrumental variables
6. regression discontinuity

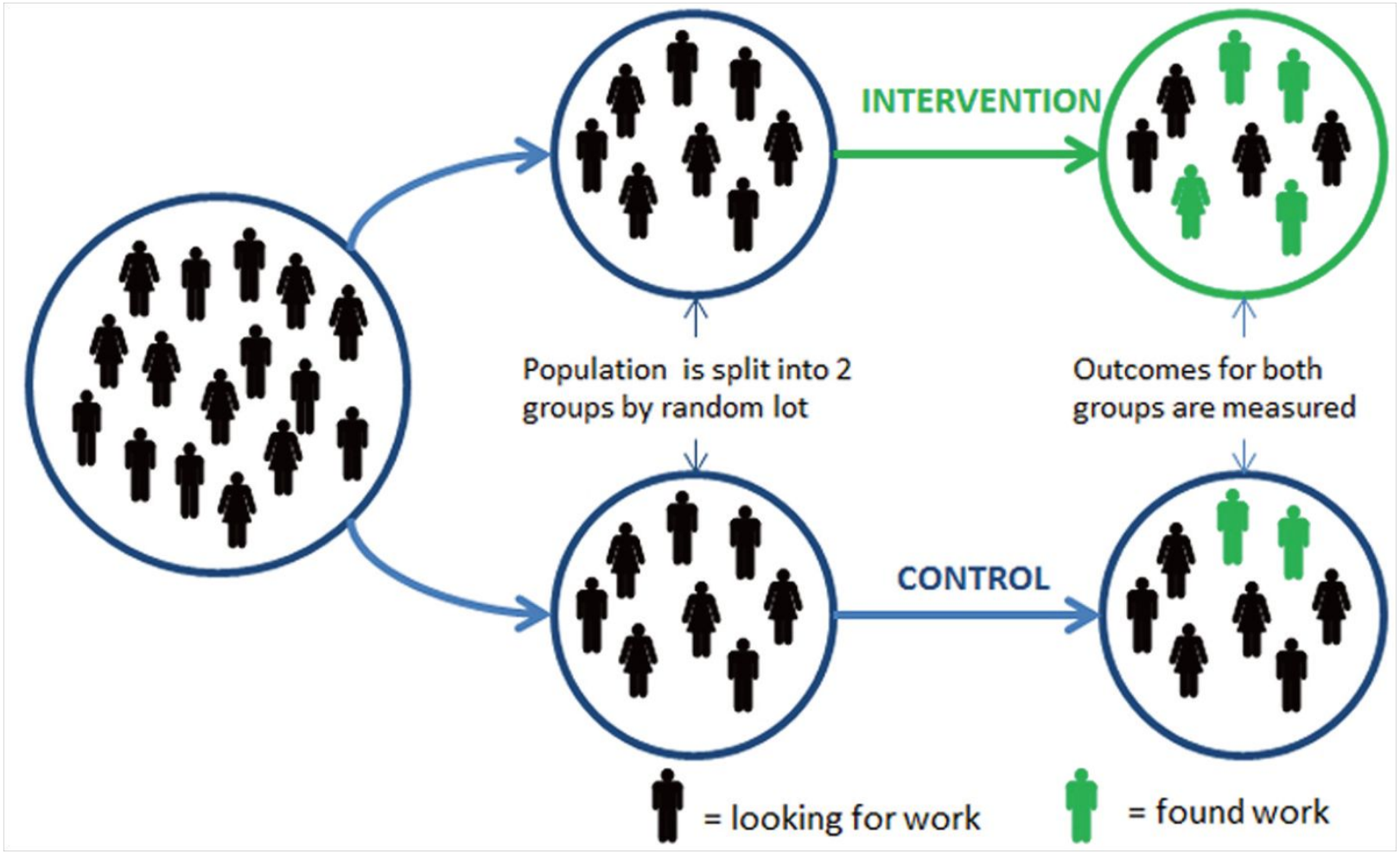
See also, “10 Strategies for figuring out if X caused Y,” EGAP Methods Guide.

causal methodologies

1. **randomized controlled trials (RCTs)**
2. "natural" experiments
3. pre-post comparisons
4. multiple regression / matching
5. instrumental variables
6. regression discontinuity

See also, "10 Strategies for figuring out if X caused Y," EGAP Methods Guide.

RCTs


INTERVENTION
Population is split into 2 groups by random lot
Outcomes for both groups are measured
CONTROL
= looking for work
= found work

body-worn camera study

bwc.thelab.dc.gov

causal methodologies

1. randomized controlled trials (RCTs)
2. **“natural” experiments**
3. pre-post comparisons
4. multiple regression / matching
5. instrumental variables
6. regression discontinuity

See also, “10 Strategies for figuring out if X caused Y,” EGAP Methods Guide.

causal methodologies

1. randomized controlled trials (RCTs)
2. “natural” experiments
3. **pre-post comparisons**
4. multiple regression / matching
5. instrumental variables
6. regression discontinuity

See also, “10 Strategies for figuring out if X caused Y,” EGAP Methods Guide.

causal methodologies

1. randomized controlled trials (RCTs)
2. “natural” experiments
3. pre-post comparisons
4. **multiple regression / matching**
5. instrumental variables
6. regression discontinuity

See also, “10 Strategies for figuring out if X caused Y,” EGAP Methods Guide.

```
Run MATRIX procedure:

HC Method
 0

Criterion Variable
 chose_AB

Model Fit:
       R-sq          F        df1        df2          p
      .0545    12.9651     9.0000  1775.0000      .0000

Heteroskedasticity-Consistent Regression Results
              Coeff     SE(HC)          t      P>|t|
Constant      .4749      .0374    12.7038      .0000
language     -.0390      .0229    -1.7050      .0884
saw_scen     -.2057      .0476    -4.3266      .0000
saw_sc_1     -.0421      .0495     -.8503      .3953
saw_sc_2      .0309      .0507      .6093      .5424
saw_sc_3      .0243      .0503      .4829      .6292
saw_sc_4     -.0159      .0500     -.3190      .7498
saw_sc_5     -.1196      .0484    -2.4727      .0135
saw_sc_6     -.0122      .0498     -.2457      .8060
saw_sc_7      .2393      .0477     5.0122      .0000

Covariance Matrix of Parameter Estimates
            Constant   language   saw_scen   saw_sc_1   saw_sc_2   saw_sc_3   saw_sc_4   saw_sc_5
Constant      .0014     -.0003     -.0013     -.0013     -.0013     -.0013     -.0013     -.0013
language     -.0003      .0005      .0000      .0000      .0000      .0000      .0000      .0000
saw_scen     -.0013      .0000      .0023      .0013      .0013      .0013      .0013      .0013
saw_sc_1     -.0013      .0000      .0013      .0025      .0013      .0013      .0013      .0013
saw_sc_2     -.0013      .0000      .0013      .0013      .0026      .0013      .0013      .0013
saw_sc_3     -.0013      .0000      .0013      .0013      .0013      .0025      .0013      .0013
saw_sc_4     -.0013      .0000      .0013      .0013      .0013      .0013      .0025      .0013
saw_sc_5     -.0013      .0000      .0013      .0013      .0013      .0013      .0013      .0023
saw_sc_6     -.0012      .0000      .0013      .0013      .0013      .0013      .0013      .0013
saw_sc_7     -.0013      .0000      .0013      .0013      .0013      .0013      .0013      .0013

------ END MATRIX -----
```

The regression was significant, (R^2 = .05), $F(9, 1775) = 12.96$, $p < .001$. But the effect of language condition was marginal and had a negative rather than the expected positive slope, β = -.039, heteroskedasticity-consistent (HC) SE = .-23, p = .088.

Regression

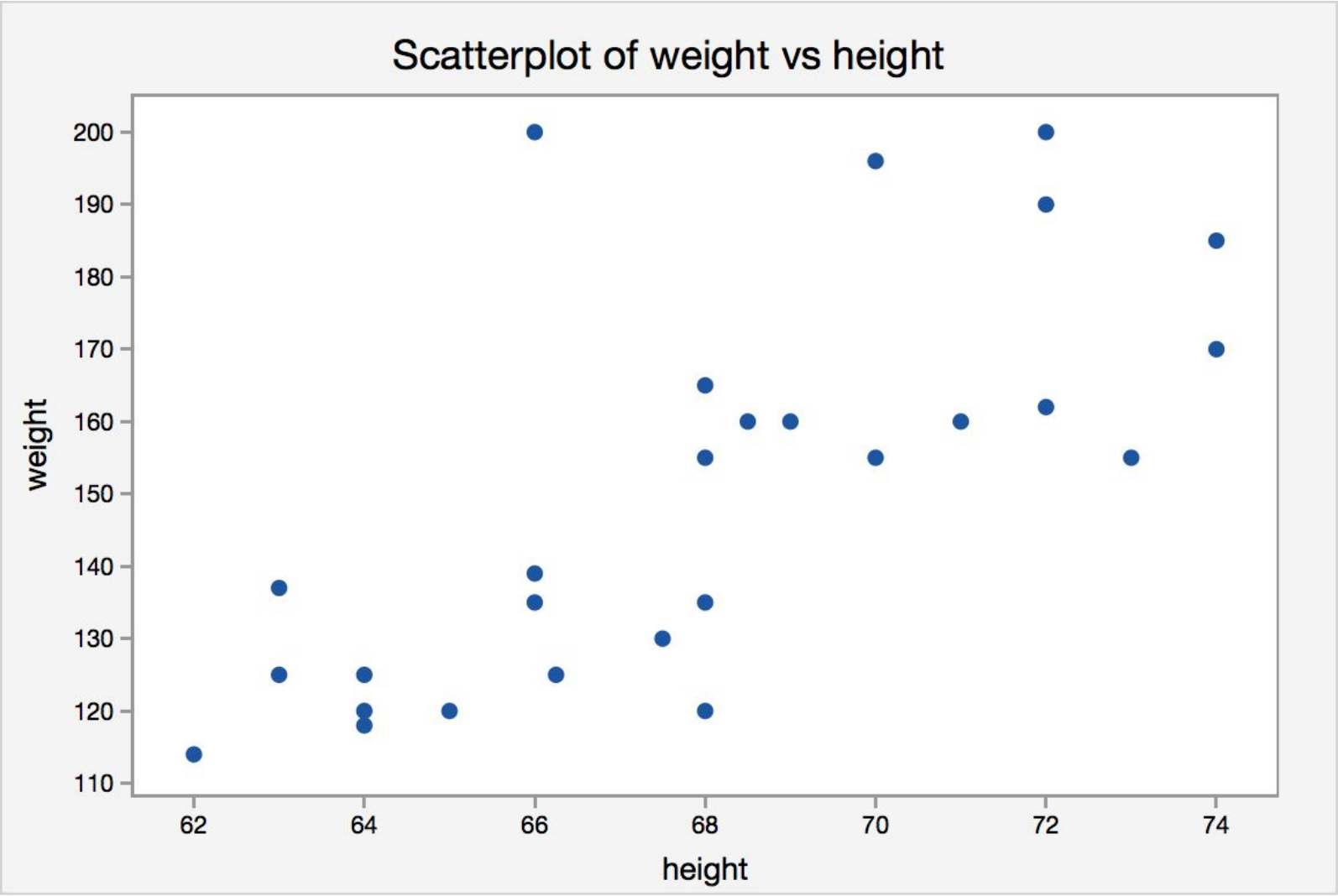
the basics

$$\hat{y}_i = a + b_1 x_1$$

- Weight = 115 + 8.6(Height_Inches)
  - all else equal, each additional inch of height predicts an additional 8.6 pounds of weight

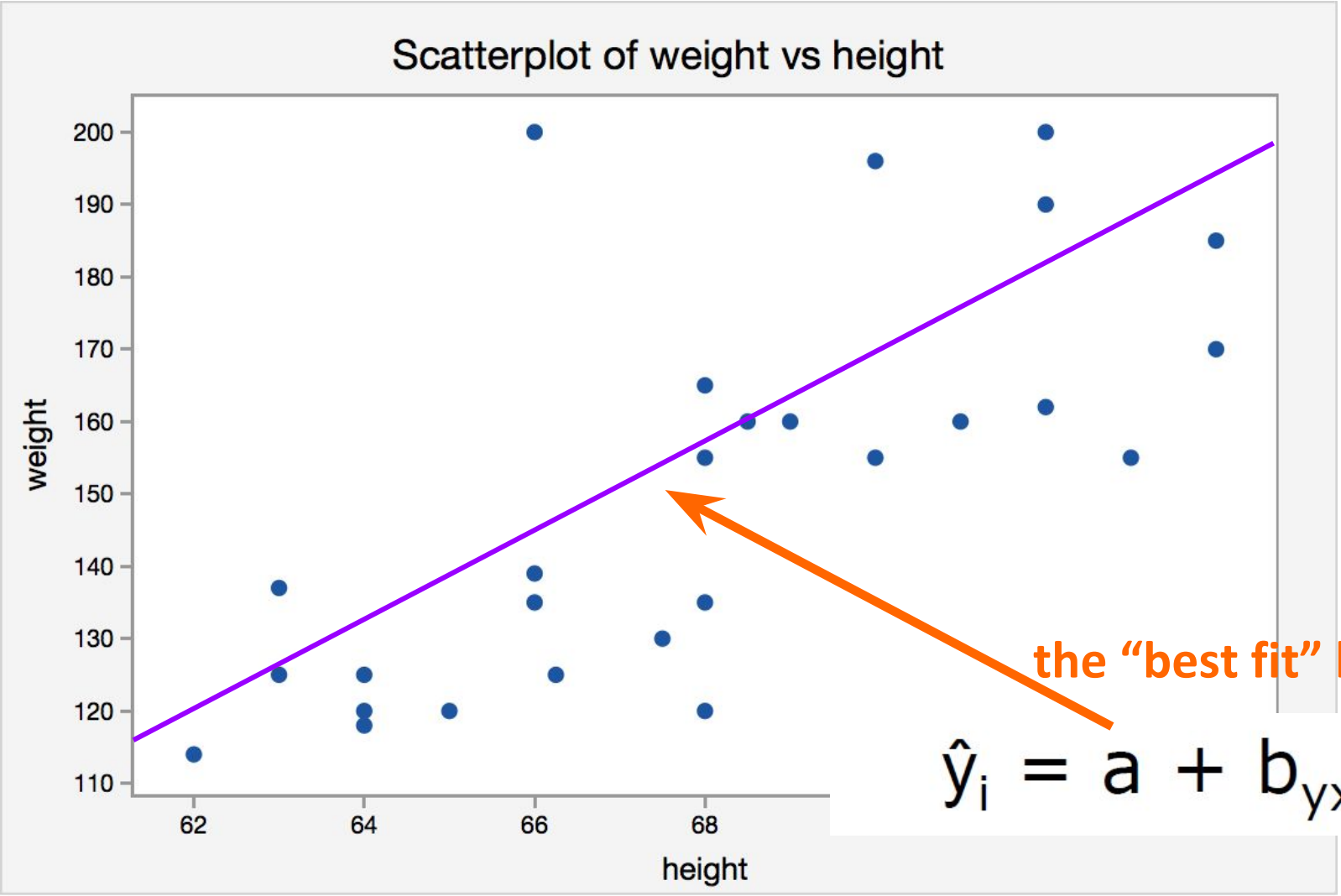
Regression

the basics



Regression

the basics

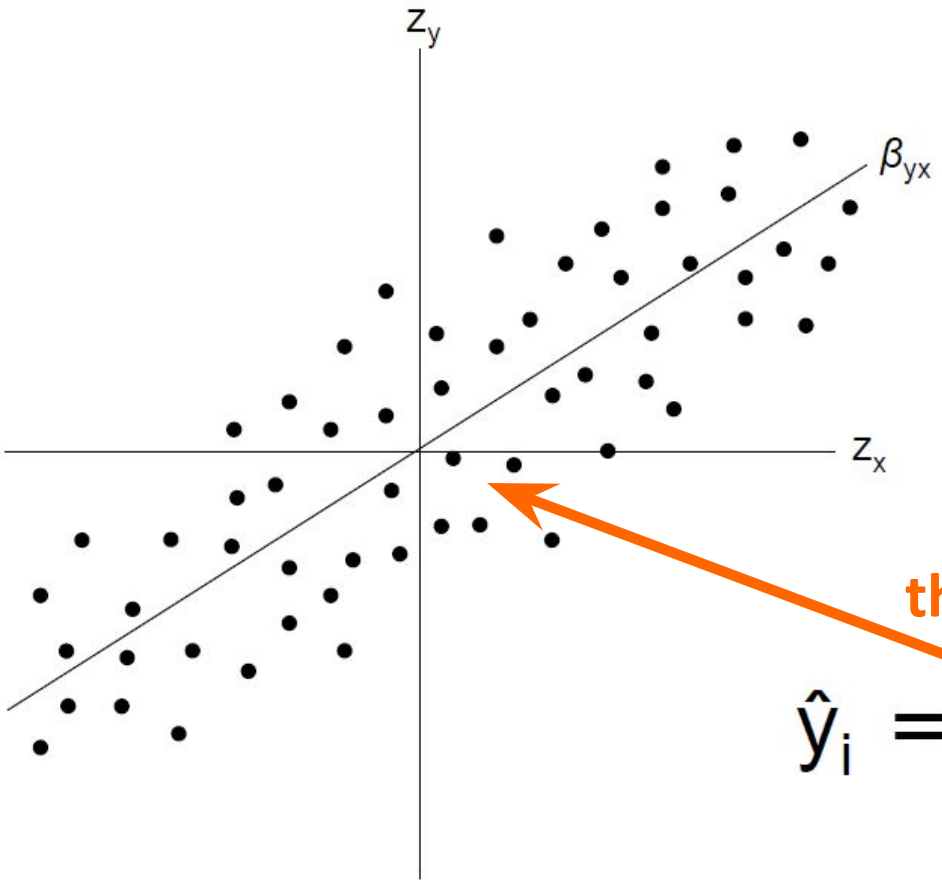


Regression

- An interactive app, to get an intuitive feel for the “best fit” line:
  - http://www.shodor.org/interactivate/activities/Regression/

Regression

visualizing one predictor



the "best fit" line

$$\hat{y}_i = a + b_{yx}x_i$$

Regression

visualizing two predictors

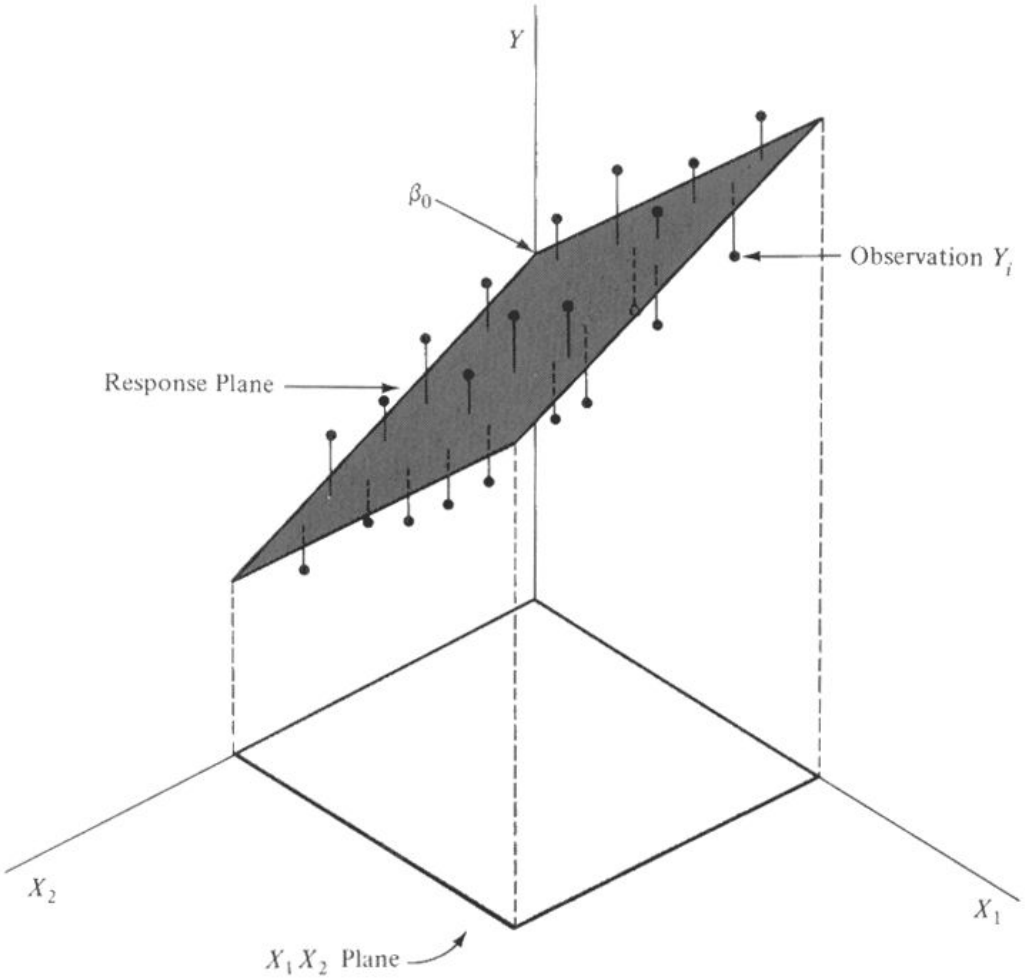


Image from: http://geography.uoregon.edu/bartlein/courses/geog495/lectures/lec13.htm

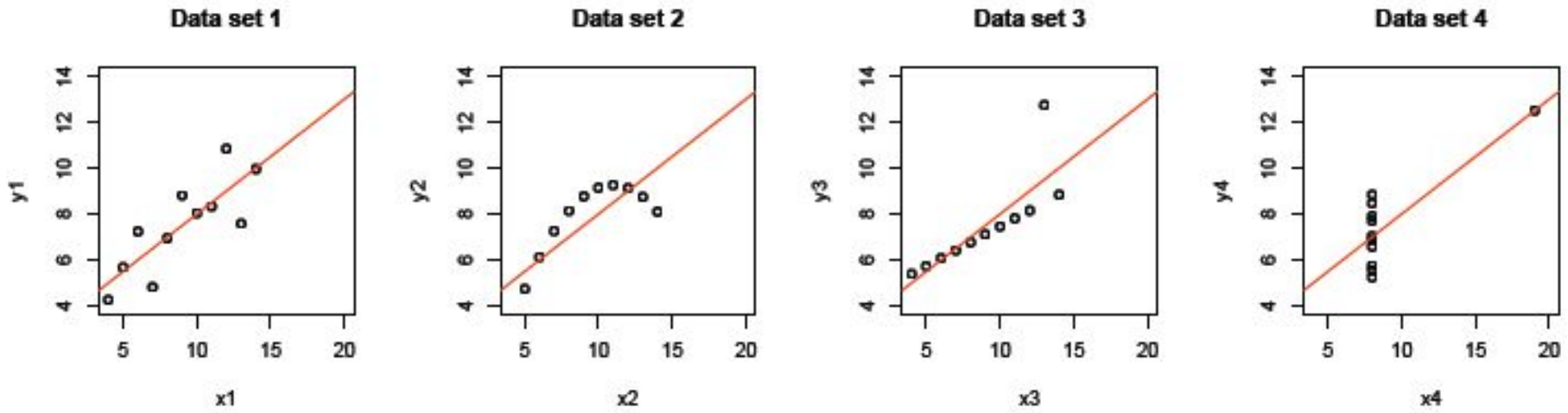
Regression

visualizing three plus predictors

Regression

two tips

1. Look at the data.



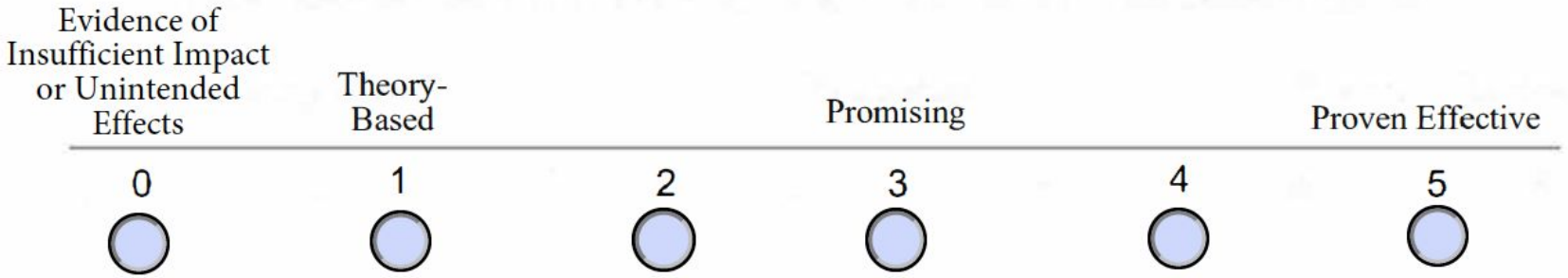
2. Ask yourself whether all the relevant predictors are included in the model.

the evidence scale

an evidence continuum

6. Evidence Scale Ranking

Please rank the proposed initiative's current level of evidentiary support on a scale from 0-5, based on the RI Evidence Scale, with one being the least evidentiary support and five being the most evidentiary support. You can use tools like the Pew Results First Clearinghouse and the Social Programs That Work database to determine whether the initiative you are proposing has been rigorously evaluated in other jurisdictions. The Office of Management & Budget understands that the majority of agency requests will likely not be in the top evidence tiers at the point of submittal, and you should certainly feel free to submit requests that are "theory-based" rather than evidence based. Please note that "theory-based" submissions should include a robust and compelling measurement and evaluation plan in the Performance Measurement section.



when and how should you generate more data?

what don't you know, that matters?

inputs	outputs	outcomes	impacts
staff facilities materials etc.	# people served # job apps submitted	# employed average wages	increase employment or wages CAUSED BY the program

implementation

outcome

impact

Forward Looking Opportunities to Develop an EvidenceBase

14. What methodologies will you use for program evaluation? Check all that apply.

- ☐ No evaluation is planned
- ☐ Measurement of the resources (e.g. staff, material expenses) required to deploy the service or initiative
- ☐ Measurement of how many people use the service or initiative
- ☐ Focus groups, surveys, or other qualitative methods that ask people about their experience with the service/initiative
- ☐ Measurement of outcomes at a point in time or over time (e.g. monthly or quarterly reports of student test scores, average wages, crime reports, park admissions, medical claims, etc.)
- ☐ Measurement of causal impact with econometric methods (e.g. regression discontinuity, multiple regression, matched controls, instrumental variables)
- ☐ Measurement of causal impact with a randomized control trial (RCT)
- ☐ Other:

what effect size(s)?

Opportunity:
low-cost RCTs / evaluation

See generally, Feldman & Haskins, "Low-cost randomized controlled trials" (2016) .

11. Which of the following best describe the type(s) of data that you currently use for evaluation? Check all that apply.

- ☐ No data is collected
- ☐ Qualitative data is collected (e.g. participant demographic information)
- ☐ Quantitative data is collected (e.g. number of participants)
- ☐ Aggregate-level data is collected (e.g. % of students who qualify for free/reduced lunch)
- ☐ Individual-level data is collected (e.g. household income amount for an individual student)

12. Which of the following best describes the source(s) of data that you currently use for evaluation? Check all that apply.

- ☐ No data is collected
- ☐ We use existing data that has been collected by another state agency, the federal government, a private entity, or another source
- ☐ We use existing data that has been collected by our agency for a different purpose
- ☐ We collect initiative-specific data

basic tips to assess data quality

1. Look at your raw data
   a. What's missing?
   b. Do the entries make sense?
   c. Do formats change over time?
2. Create a data dictionary
3. Do field work to understand how data were generated.
   a. Pro Tip: Map it to your process map
4. Look at visualizations of your data
5. Compute basic descriptive statistics

partnership opportunities

18. Have you identified research partners to help evaluate the initiative? If so, please describe who. If not, what type of research partnership, if any, would be helpful?

19. Do you want to speak with a methods expert for a consultation on what evaluation methods might be best for learning about and optimizing the performance of your initiative?

- ❑ No, we do not need a methods expert
- ❑ No, we have already engaged a methods expert
- ❑ Maybe, it might be useful, we're open to it
- ❑ Yes, that would be helpful

resources

- Budget Instructions
- Decision Package Template
- Examples (good and bad)
- Website Trainer
- **Office Hours**

- *Your feedback, much appreciated!*

discussion

connect with The Policy Lab

- Explore at thepolicylab.brown.edu, and sign up for our listserv for updates, events, and ways to collaborate.
- Check out the podcast at thirtythousandleagues.com.