

# Measuring Protest Participation as a Latent Variable

Frederick Solt  
University of Iowa

# Two Forerunning Literatures

# Two Forerunning Literatures

- Single Country Public Opinion
- Estimate latent public opinion in one country from many surveys

# Two Forerunning Literatures

- Single Country Public Opinion
- Estimate latent public opinion in one country from many surveys
  - Policy Mood: Stimson (1991)

# Two Forerunning Literatures

- Single Country Public Opinion
- Estimate latent public opinion in one country from many surveys
  - Policy Mood: Stimson (1991)
  - McGann (2014); Caughley and Warshaw (2015)

# Two Forerunning Literatures

- Single Country Public Opinion
- Estimate latent public opinion in one country from many surveys
  - Policy Mood: Stimson (1991)
  - McGann (2014); Caughley and Warshaw (2015)
- Pooled Time Series of Other Concepts
- Estimate TSCS latent variables from many cross-national time series

# Two Forerunning Literatures

- Single Country Public Opinion
  - Estimate latent public opinion in one country from many surveys
    - Policy Mood: Stimson (1991)
    - McGann (2014); Caughley and Warshaw (2015)
- Pooled Time Series of Other Concepts
  - Estimate TSCS latent variables from many cross-national time series
    - Democracy: Treier and Jackman (2008); Pemstein, Meserve, and Melton (2010)

# Two Forerunning Literatures

- Single Country Public Opinion
  - Policy Mood: Stimson (1991)
  - McGann (2014); Caughley and Warshaw (2015)
- Pooled Time Series of Other Concepts
  - Estimate TSCS latent variables from many cross-national time series
  - Democracy: Treier and Jackman (2008); Pemstein, Meserve, and Melton (2010)
  - Judicial Independence: Linzer and Staton (2013)



# Two Forerunning Literatures

- Single Country Public Opinion
  - Policy Mood: Stimson (1991)
  - McGann (2014); Caughley and Warshaw (2015)
- Pooled Time Series of Other Concepts
  - Democracy: Treier and Jackman (2008); Pemstein, Meserve, and Melton (2010)
  - Judicial Independence: Linzer and Staton (2013)
- Estimate latent public opinion in one country from many surveys
- Estimate TSCS latent variables from many cross-national time series



how to model  
incomparable survey data:  
Item Response Theory

# Two Forerunning Literatures

- Single Country Public Opinion
  - Policy Mood: Stimson (1991)
  - McGann (2014); Caughley and Warshaw (2015)
- Pooled Time Series of Other Concepts
  - Democracy: Treier and Jackman (2008); Pemstein, Meserve, and Melton (2010)
  - Judicial Independence: Linzer and Staton (2013)



how to model  
incomparable survey data:  
Item Response Theory



how to deal with sparse  
pooled cross-sections:  
Random Walk Priors

# Item Response Theory

# Item Response Theory

Responses depend on both the person ( $i$ ) and item ( $q$ ):

# Item Response Theory

Responses depend on both the person ( $i$ ) and item ( $q$ ):

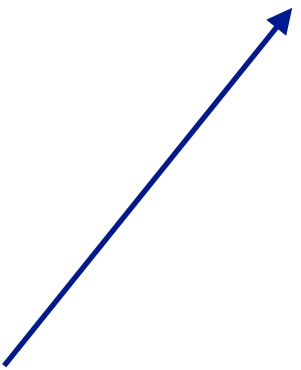
$$Pr(y_{iq} = 1) = \text{Logit}^{-1}((\alpha_i - \beta_q)/\gamma_q)$$

# Item Response Theory

Responses depend on both the person ( $i$ ) and item ( $q$ ):

$$Pr(y_{iq} = 1) = \text{Logit}^{-1}((\alpha_i - \beta_q)/\gamma_q)$$

characteristic  
(of person  $i$ )

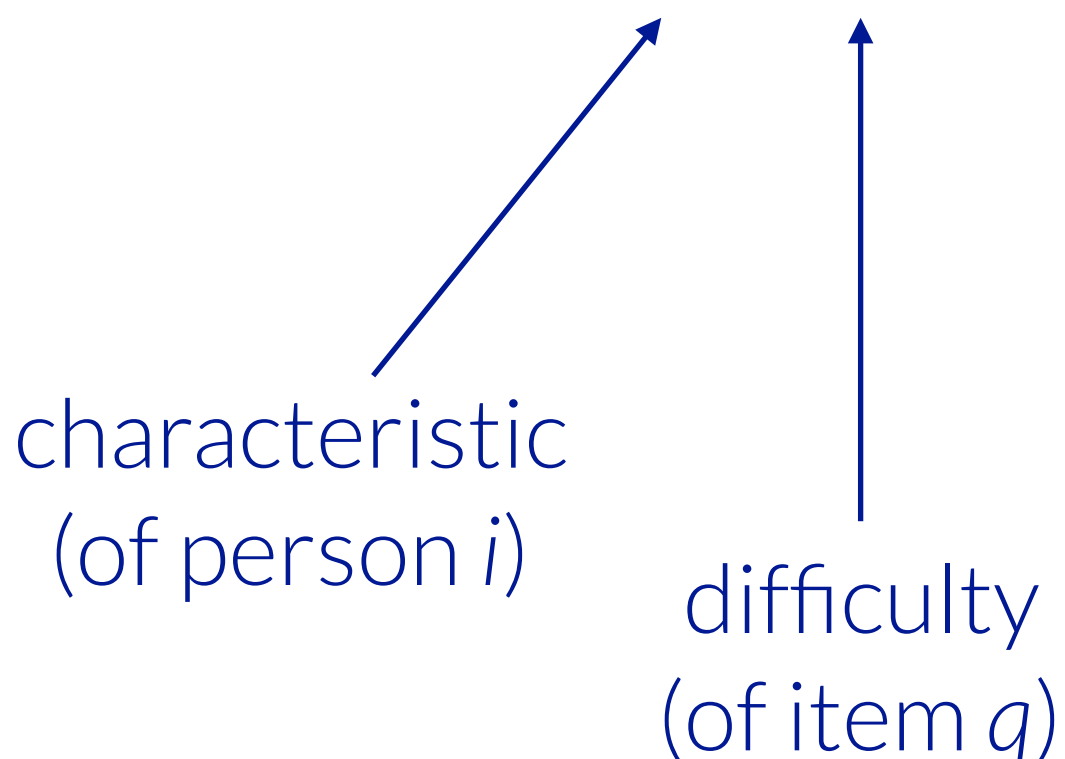


# Item Response Theory

Responses depend on both the person ( $i$ ) and item ( $q$ ):

$$Pr(y_{iq} = 1) = \text{Logit}^{-1}((\alpha_i - \beta_q)/\gamma_q)$$

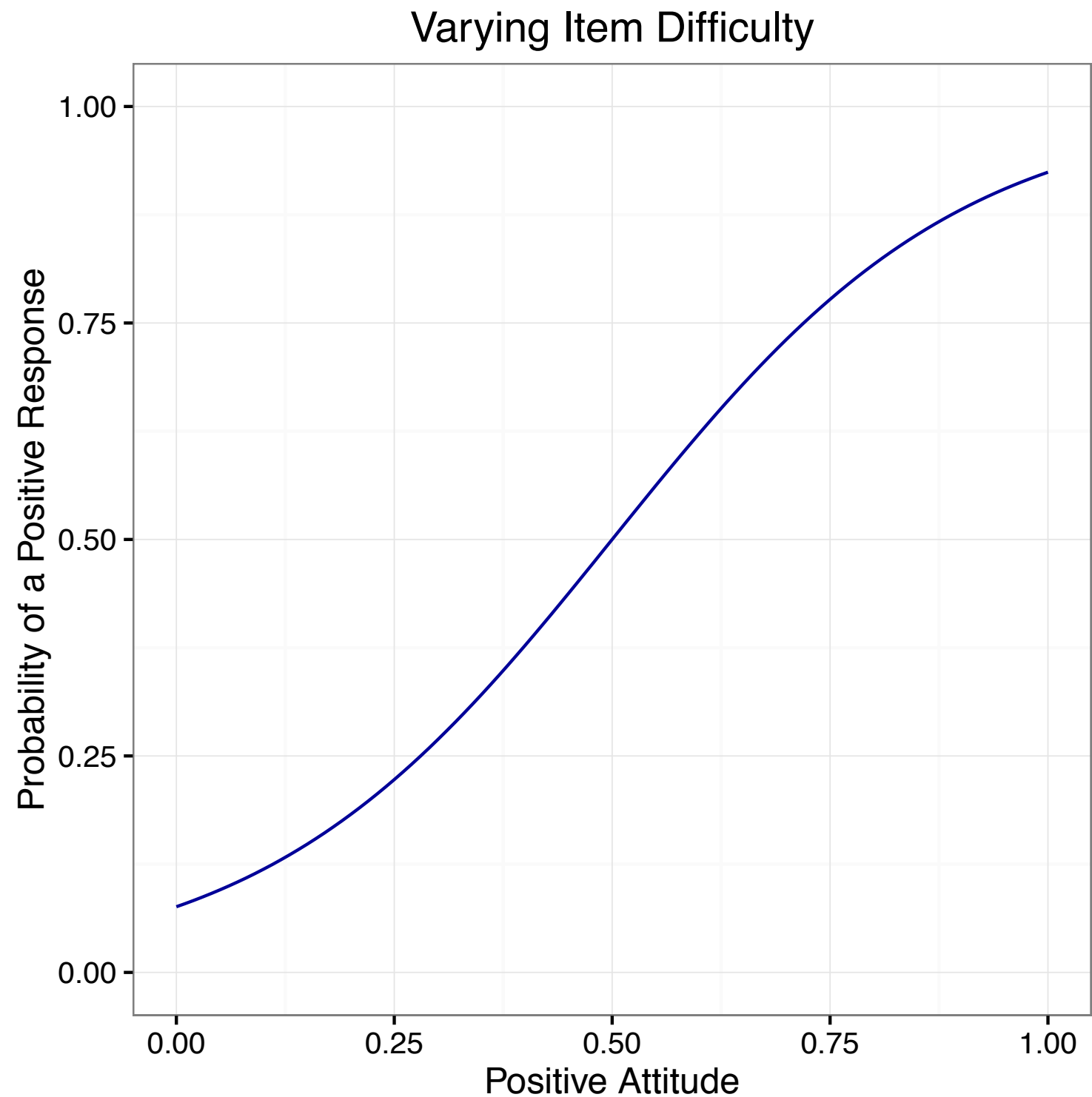
characteristic  
(of person  $i$ )



The diagram consists of two blue arrows. One arrow originates from the text 'characteristic (of person i)' and points diagonally upwards and to the right, terminating at the  $\alpha_i$  term in the equation above. The second arrow originates from the text 'difficulty (of item q)' and points diagonally upwards and to the left, terminating at the  $\beta_q$  term in the equation above.

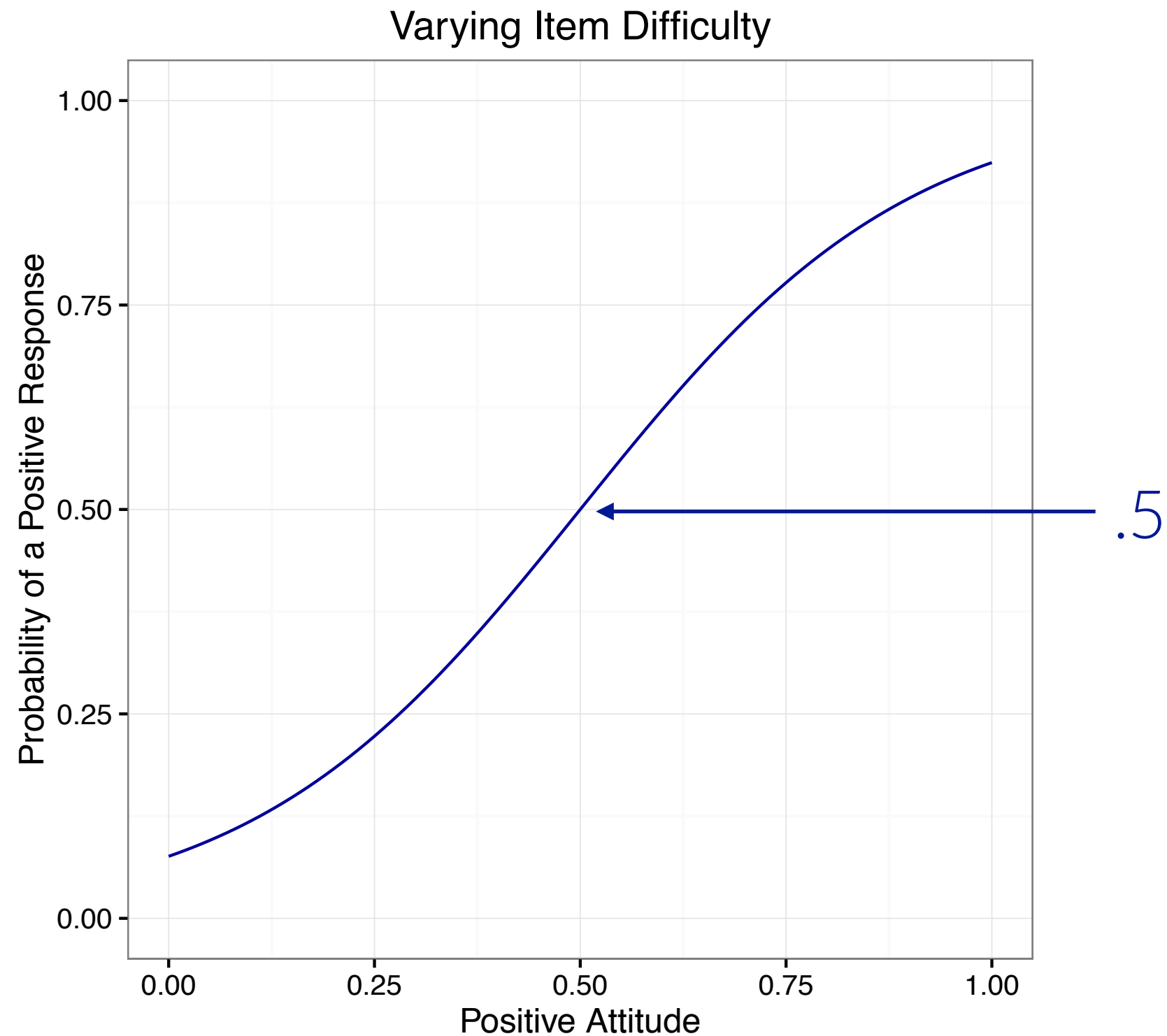
difficulty  
(of item  $q$ )

# Item Response Theory

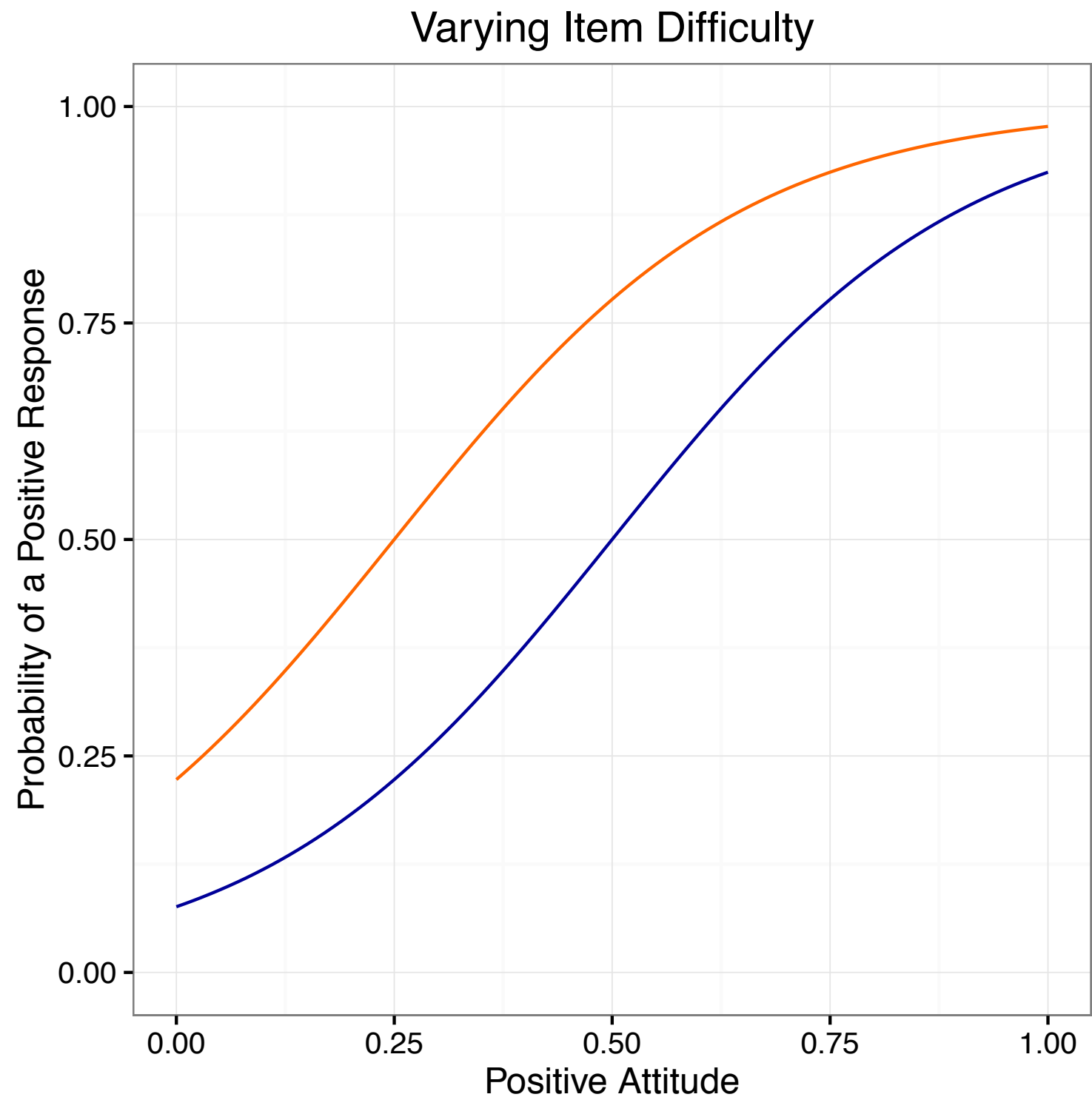




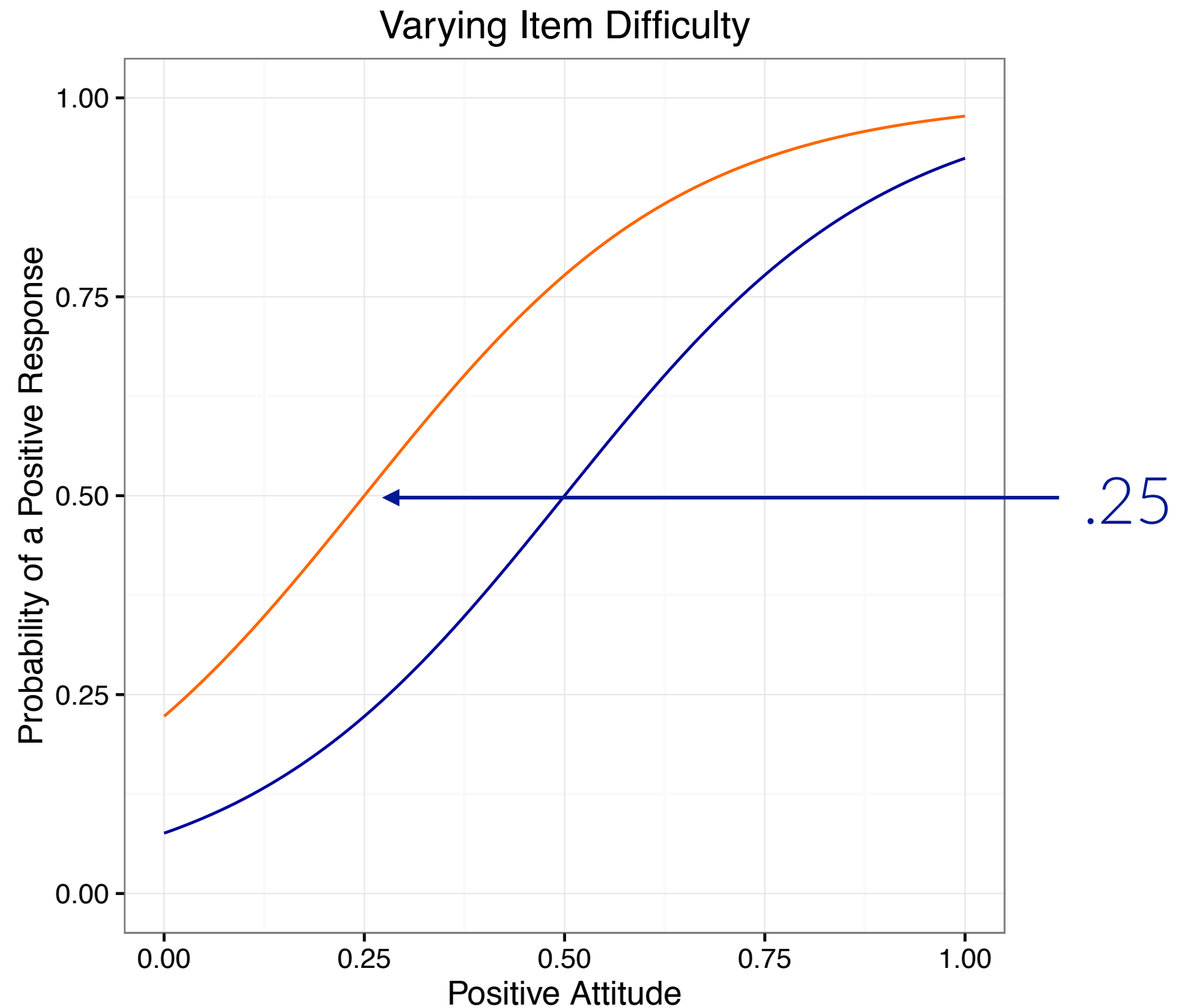
# Item Response Theory



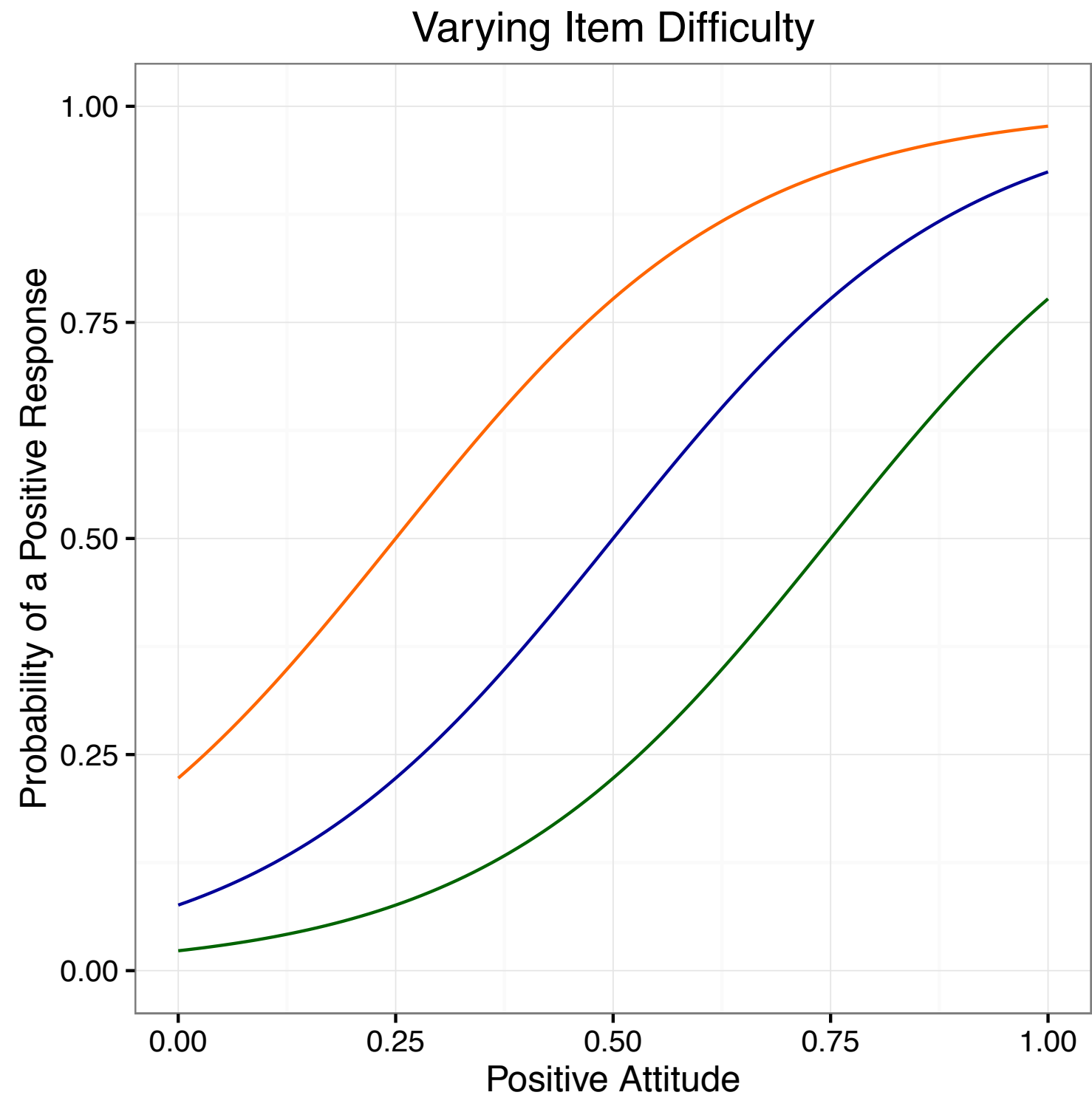
# Item Response Theory



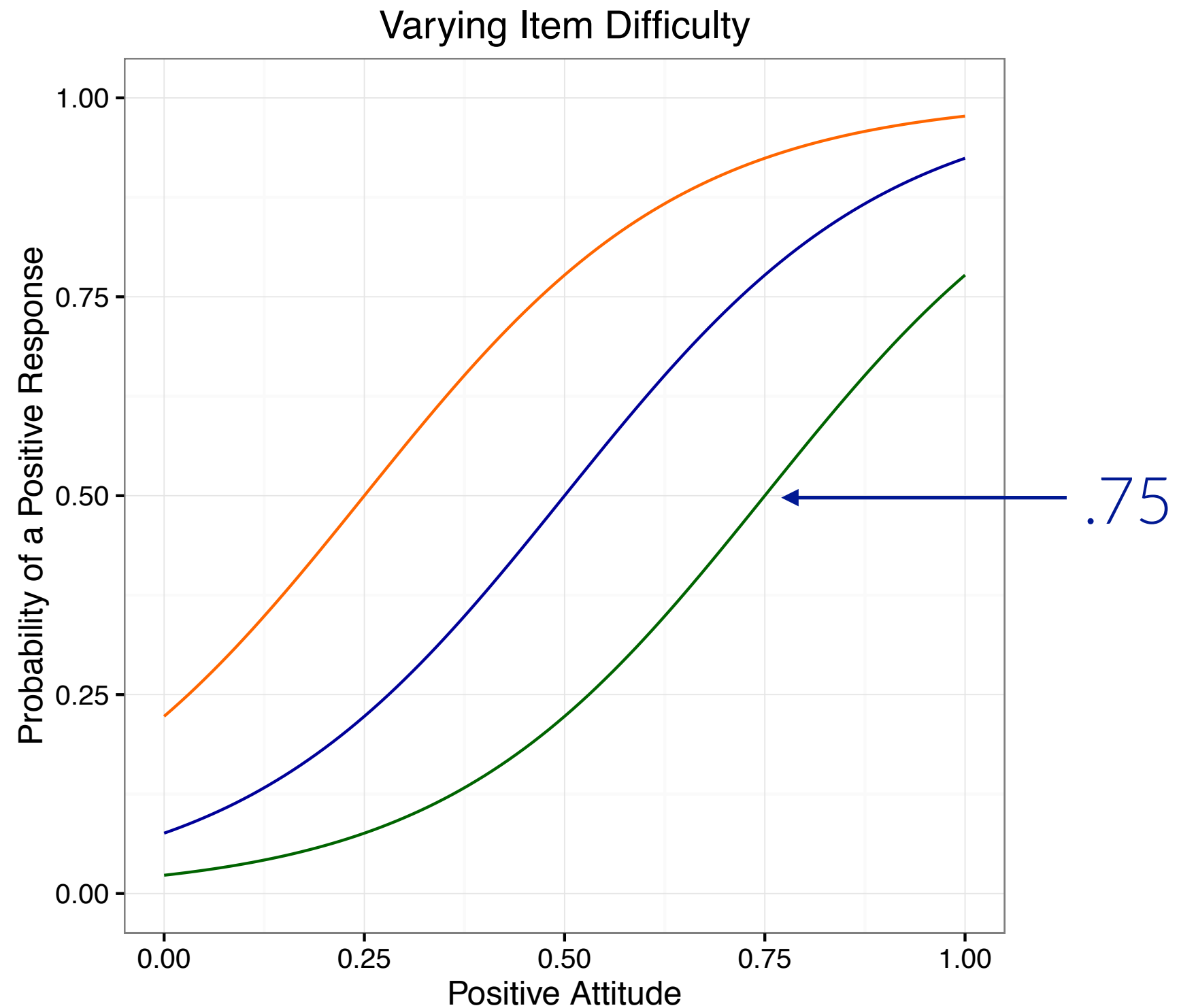
# Item Response Theory



# Item Response Theory



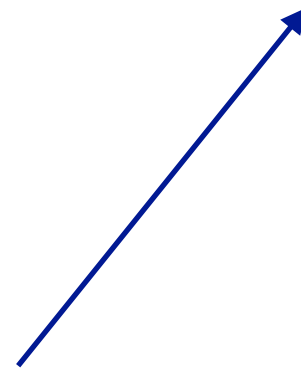
# Item Response Theory



# Item Response Theory

Responses depend on both the person ( $i$ ) and item ( $q$ ):

$$Pr(y_{iq} = 1) = \text{Logit}^{-1}((\alpha_i - \beta_q)/\gamma_q)$$



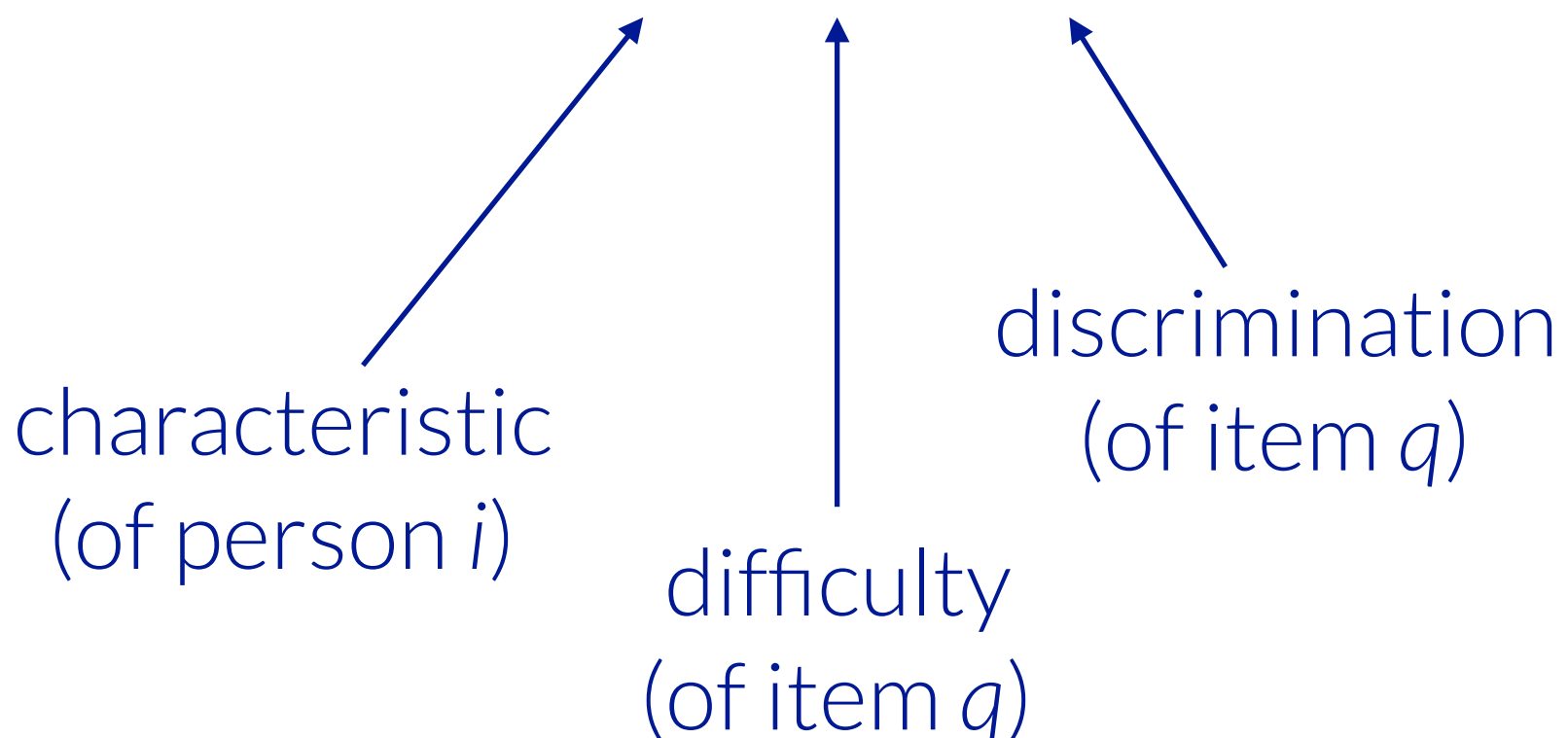
difficulty  
(of item  $q$ )

# Item Response Theory

Responses depend on both the person ( $i$ ) and item ( $q$ ):

$$Pr(y_{iq} = 1) = \text{Logit}^{-1}((\alpha_i - \beta_q)/\gamma_q)$$

characteristic  
(of person  $i$ )

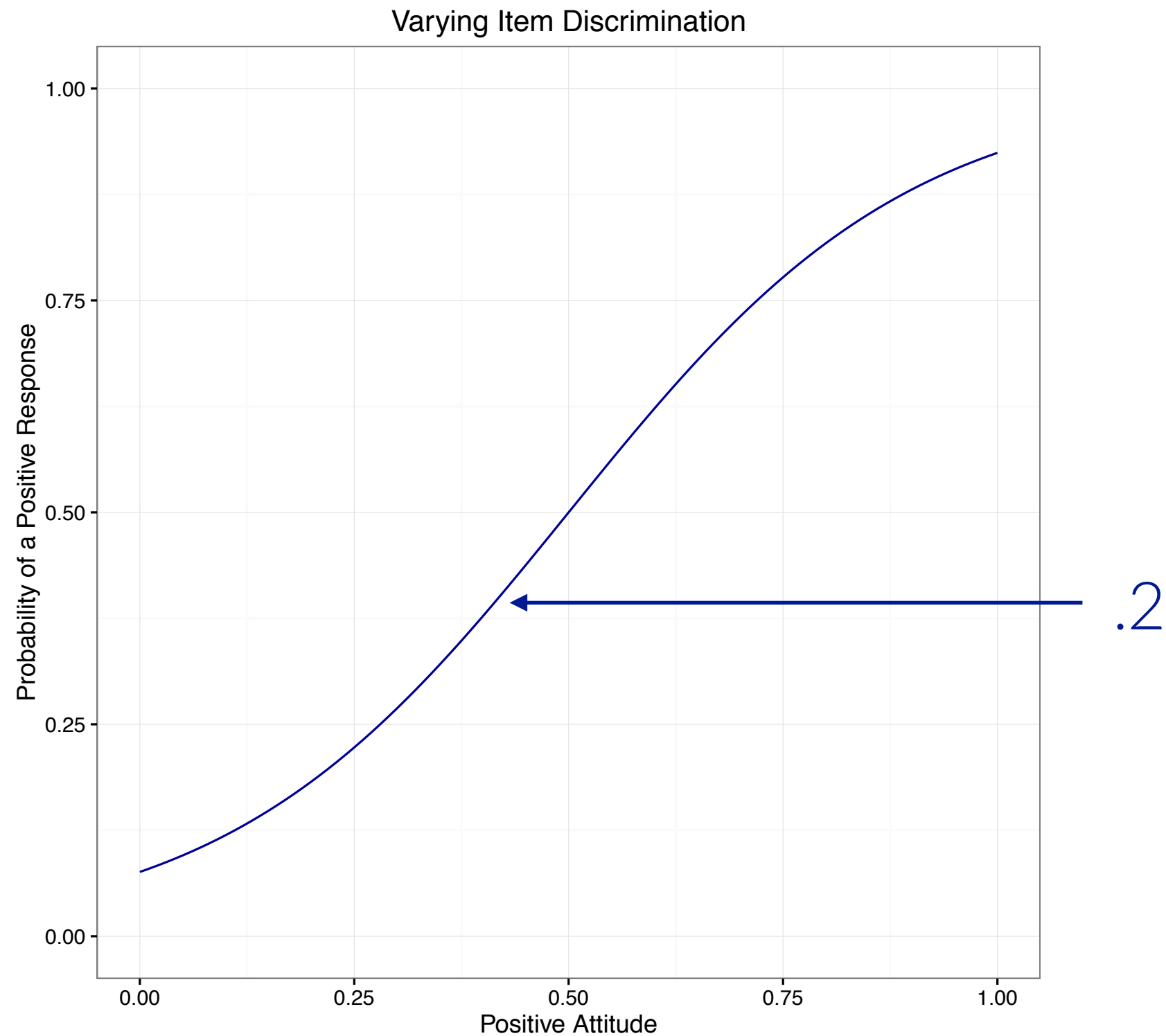


The diagram consists of three blue arrows pointing upwards towards the equation. The first arrow originates from the text 'characteristic (of person i)' and points to the parameter  $\alpha_i$ . The second arrow originates from the text 'difficulty (of item q)' and points to the parameter  $\beta_q$ . The third arrow originates from the text 'discrimination (of item q)' and points to the parameter  $\gamma_q$ .

difficulty  
(of item  $q$ )

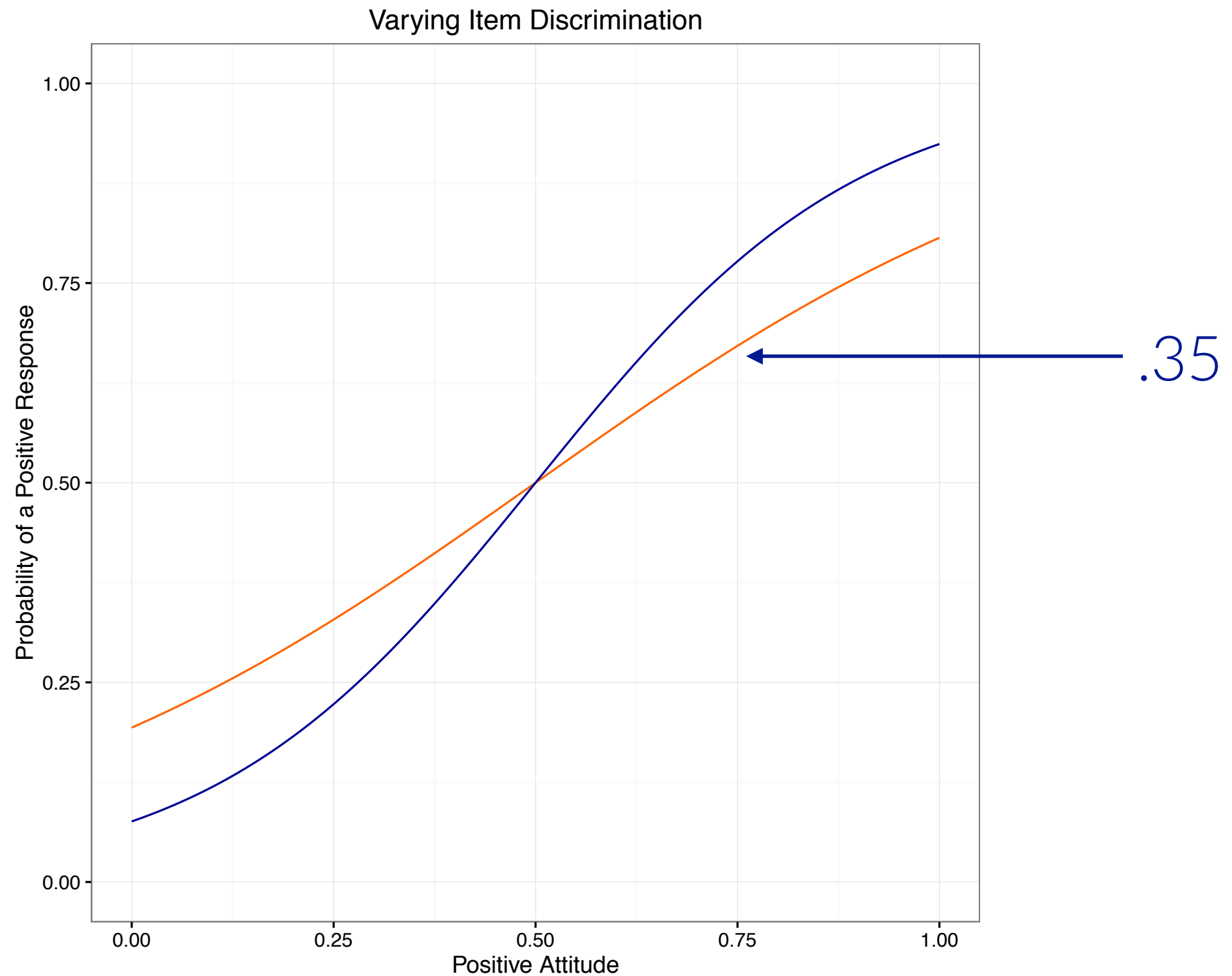
discrimination  
(of item  $q$ )

# Item Response Theory

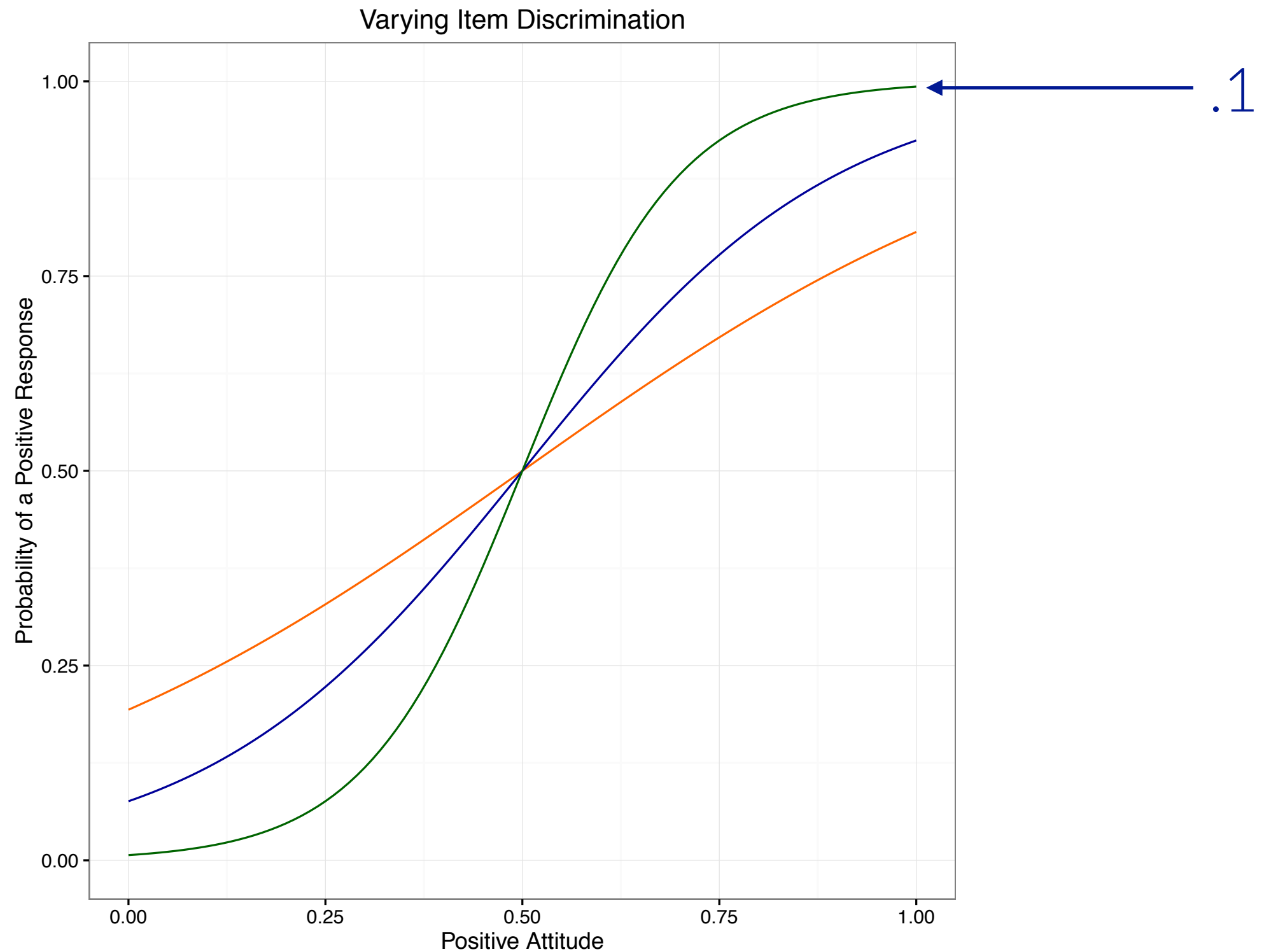




# Item Response Theory



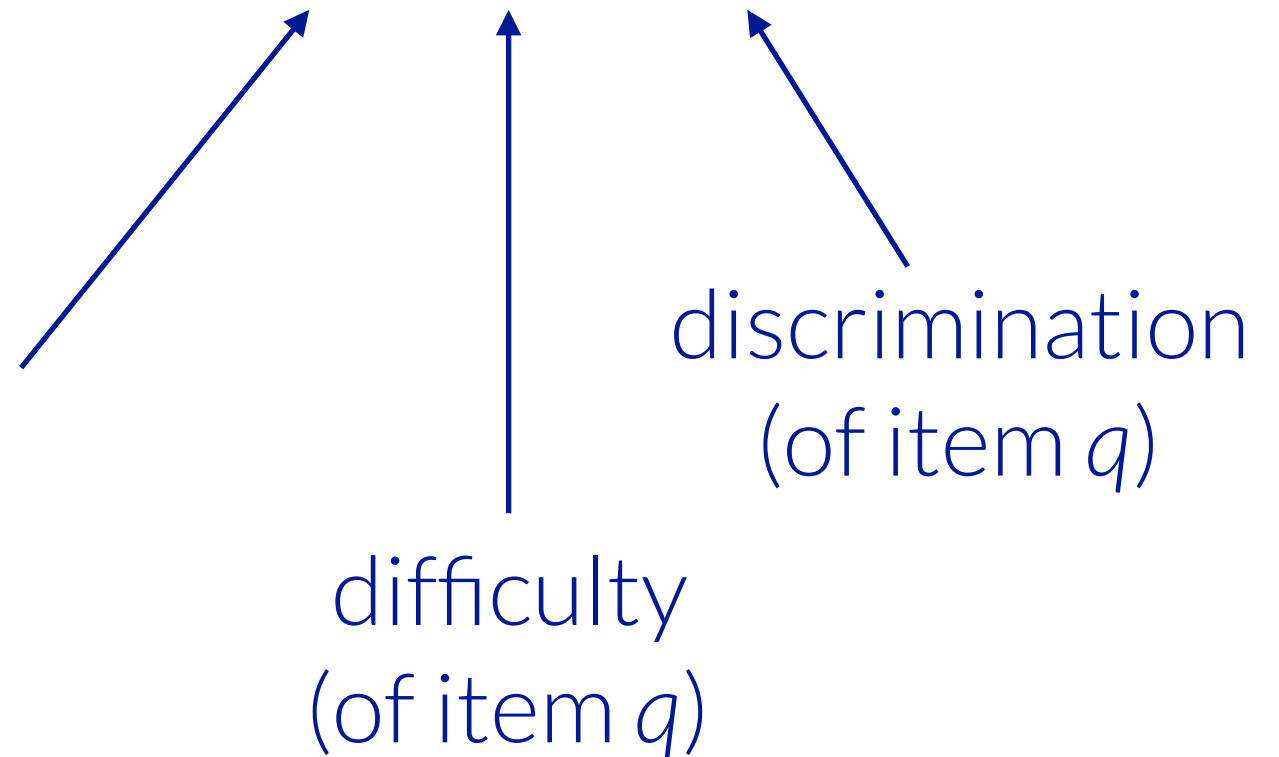
# Item Response Theory



# Item Response Theory

Responses depend on both the person ( $i$ ) and item ( $q$ ):

$$Pr(y_{iq} = 1) = \text{Logit}^{-1}((\alpha_i - \beta_q)/\gamma_q)$$



# Item Response Theory

Responses depend on both the person ( $i$ ) and item ( $q$ ):

$$Pr(y_{iq} = 1) = \text{Logit}^{-1}((\alpha_i - \beta_q)/\gamma_q)$$

characteristic  
(of person  $i$ )



The diagram consists of three blue arrows pointing upwards towards the equation. The first arrow originates from the text 'characteristic (of person i)' and points to the parameter  $\alpha_i$ . The second arrow originates from the text 'difficulty (of item q)' and points to the parameter  $\beta_q$ . The third arrow originates from the text 'discrimination (of item q)' and points to the parameter  $\gamma_q$ .

difficulty  
(of item  $q$ )

discrimination  
(of item  $q$ )

# Item Response Theory

scaled to country-year level

Responses depend on the country ( $k$ ), year ( $t$ ), and item ( $q$ ):

$$m_{ktq} = \text{Logit}^{-1}((\bar{\alpha}_{kt} - \beta_q) / \gamma_q)$$

mean  
characteristic  
(in country  $k$   
at time  $t$ )



difficulty  
(of item  $q$ )

discrimination  
(of item  $q$ )

# Random Walk Priors

for country-year parameters

Values of time-varying parameters next year ( $t+1$ ) in country  $k$  depend on their values in this year ( $t$ ):

$$\bar{\alpha}_{kt+1} = \text{Normal}(\bar{\alpha}_{kt}, \sigma_k^2)$$

# Random Walk Priors

for country-year parameters

Values of time-varying parameters next year ( $t+1$ ) in country  $k$  depend on their values in this year ( $t$ ):

$$\bar{\alpha}_{kt+1} = \text{Normal}(\bar{\alpha}_{kt}, \sigma_k^2)$$

this year's value  
(in country  $k$ )



# Random Walk Priors

for country-year parameters

Values of time-varying parameters next year ( $t+1$ ) in country  $k$  depend on their values in this year ( $t$ ):

$$\bar{\alpha}_{kt+1} = \text{Normal}(\bar{\alpha}_{kt}, \sigma_k^2)$$

this year's value  
(in country  $k$ )



variance in values  
over time  
(in country  $k$ )





# Random Walk Priors

for country-year parameters

Values of time-varying parameters next year ( $t+1$ ) in country  $k$  depend on their values in this year ( $t$ ):

$$\bar{\alpha}_{kt+1} = \text{Normal}(\bar{\alpha}_{kt}, \sigma_k^2)$$

next year's value  
(in country  $k$ )

this year's value  
(in country  $k$ )

variance in values  
over time  
(in country  $k$ )

# Data

# Data

- Collect cross-national and national surveys



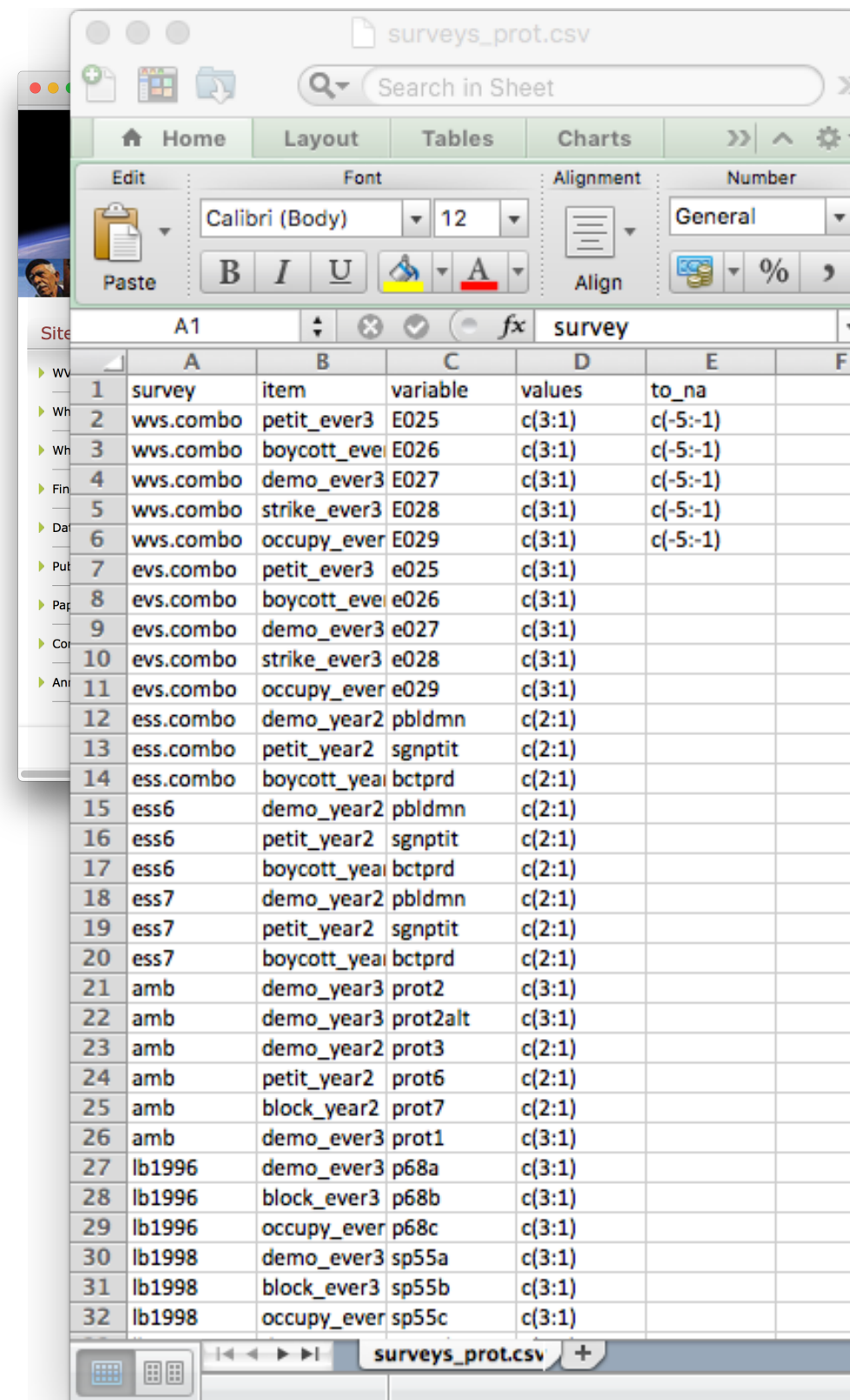
# Data

- Collect cross-national and national surveys
  - Standardize country, year, and weight variables



# Data

- Collect cross-national and national surveys
  - Standardize country, year, and weight variables
- Identify and classify relevant survey items



The screenshot shows a spreadsheet application window titled "surveys\_prot.csv". The spreadsheet contains a table with the following data:

	A	B	C	D	E	F
1	survey	item	variable	values	to_na	
2	wvs.combo	petit_ever3	E025	c(3:1)	c(-5:-1)	
3	wvs.combo	boycott_ever3	E026	c(3:1)	c(-5:-1)	
4	wvs.combo	demo_ever3	E027	c(3:1)	c(-5:-1)	
5	wvs.combo	strike_ever3	E028	c(3:1)	c(-5:-1)	
6	wvs.combo	occupy_ever3	E029	c(3:1)	c(-5:-1)	
7	evs.combo	petit_ever3	e025	c(3:1)		
8	evs.combo	boycott_ever3	e026	c(3:1)		
9	evs.combo	demo_ever3	e027	c(3:1)		
10	evs.combo	strike_ever3	e028	c(3:1)		
11	evs.combo	occupy_ever3	e029	c(3:1)		
12	ess.combo	demo_year2	pbldmn	c(2:1)		
13	ess.combo	petit_year2	sgnptit	c(2:1)		
14	ess.combo	boycott_year2	bctprd	c(2:1)		
15	ess6	demo_year2	pbldmn	c(2:1)		
16	ess6	petit_year2	sgnptit	c(2:1)		
17	ess6	boycott_year2	bctprd	c(2:1)		
18	ess7	demo_year2	pbldmn	c(2:1)		
19	ess7	petit_year2	sgnptit	c(2:1)		
20	ess7	boycott_year2	bctprd	c(2:1)		
21	amb	demo_year3	prot2	c(3:1)		
22	amb	demo_year3	prot2alt	c(3:1)		
23	amb	demo_year2	prot3	c(2:1)		
24	amb	petit_year2	prot6	c(2:1)		
25	amb	block_year2	prot7	c(2:1)		
26	amb	demo_ever3	prot1	c(3:1)		
27	lb1996	demo_ever3	p68a	c(3:1)		
28	lb1996	block_ever3	p68b	c(3:1)		
29	lb1996	occupy_ever3	p68c	c(3:1)		
30	lb1998	demo_ever3	sp55a	c(3:1)		
31	lb1998	block_ever3	sp55b	c(3:1)		
32	lb1998	occupy_ever3	sp55c	c(3:1)		

# Data

- Collect cross-national and national surveys
  - Standardize country, year, and weight variables
- Identify and classify relevant survey items
- Loop through items, saving the number of respondents who claimed to participate

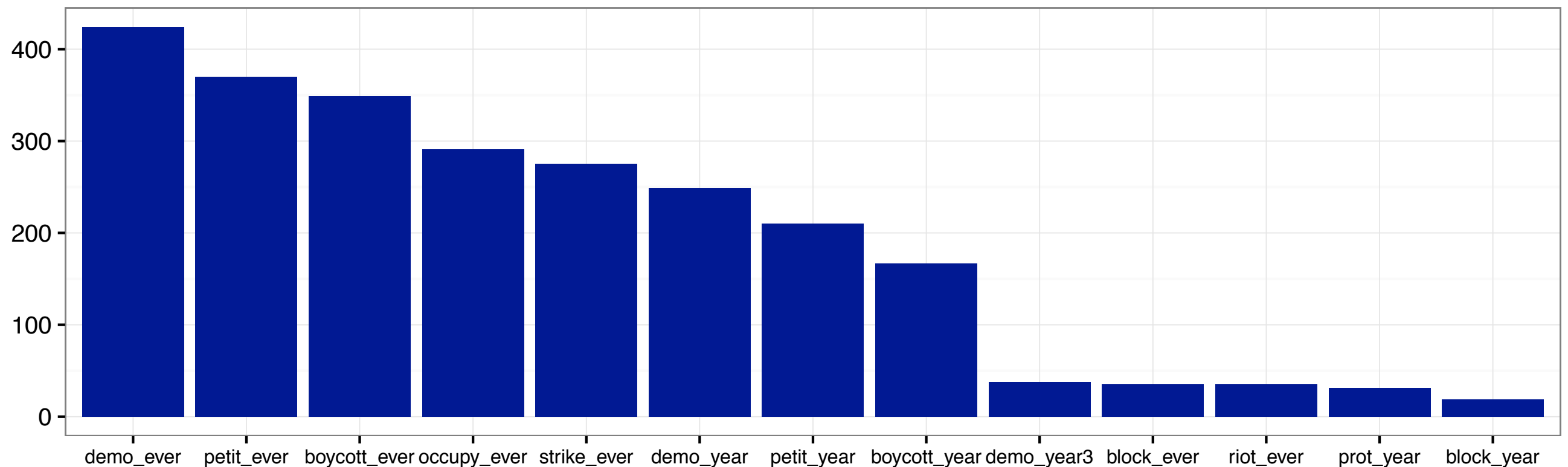
The image shows two overlapping spreadsheet windows. The left window displays a list of survey items with columns for 'survey', 'item', and 'weight'. The right window, titled 'prot\_all\_data.csv', shows a standardized dataset with columns for 'country', 'year', 'variable', 'y\_r', and 'n'. The data in the right window is as follows:

	A	B	C	F	G
	country	year	variable	y_r	n
1	Spain	1981	petit_ever3	472	1992
2	Spain	1981	boycott_ever3	177	1989
3	Spain	1981	demo_ever3	485	2032
4	Spain	1981	strike_ever3	142	2032
5	Spain	1981	occupy_ever3	61	2047
6	Spain	1990	petit_ever3	237	1268
7	Spain	1990	petit_ever3	471	2141
8	Spain	1990	boycott_ever3	103	2140
9	Spain	1990	boycott_ever3	85	1265
10	Spain	1990	demo_ever3	498	2288
11	Spain	1990	demo_ever3	345	1339
12	Spain	1990	strike_ever3	120	2216
13	Spain	1990	strike_ever3	114	1306
14	Spain	1990	occupy_ever3	54	2174
15	Spain	1990	occupy_ever3	46	1286
16	Spain	1995	petit_ever3	248	1126
17	Spain	1995	boycott_ever3	56	1109
18	Spain	1995	demo_ever3	252	1173
19	Spain	1995	strike_ever3	58	1147
20	Spain	1995	occupy_ever3	18	1130
21	Spain	1996	demo_ever3	5021	20078
22	Spain	1996	occupy_ever3	608	19705
23	Spain	1996	block_ever3	1361	19823
24	Spain	1999	petit_ever3	294	1028
25	Spain	1999	boycott_ever3	56	999
26	Spain	1999	demo_ever3	289	1074
27	Spain	1999	strike_ever3	90	1038
28	Spain	1999	occupy_ever3	32	1035
29	Spain	2000	petit_ever3	285	1077
30	Spain	2000	boycott_ever3	63	1066
31	Spain	2000	demo_ever3	296	1125

# Data

- Many surveys ask respondents' protest participation
  - WVS, EVS, ESS, LB, AmB, EQLS, etc.
- 13 different items; 2493 unique country-year-items *ktq*

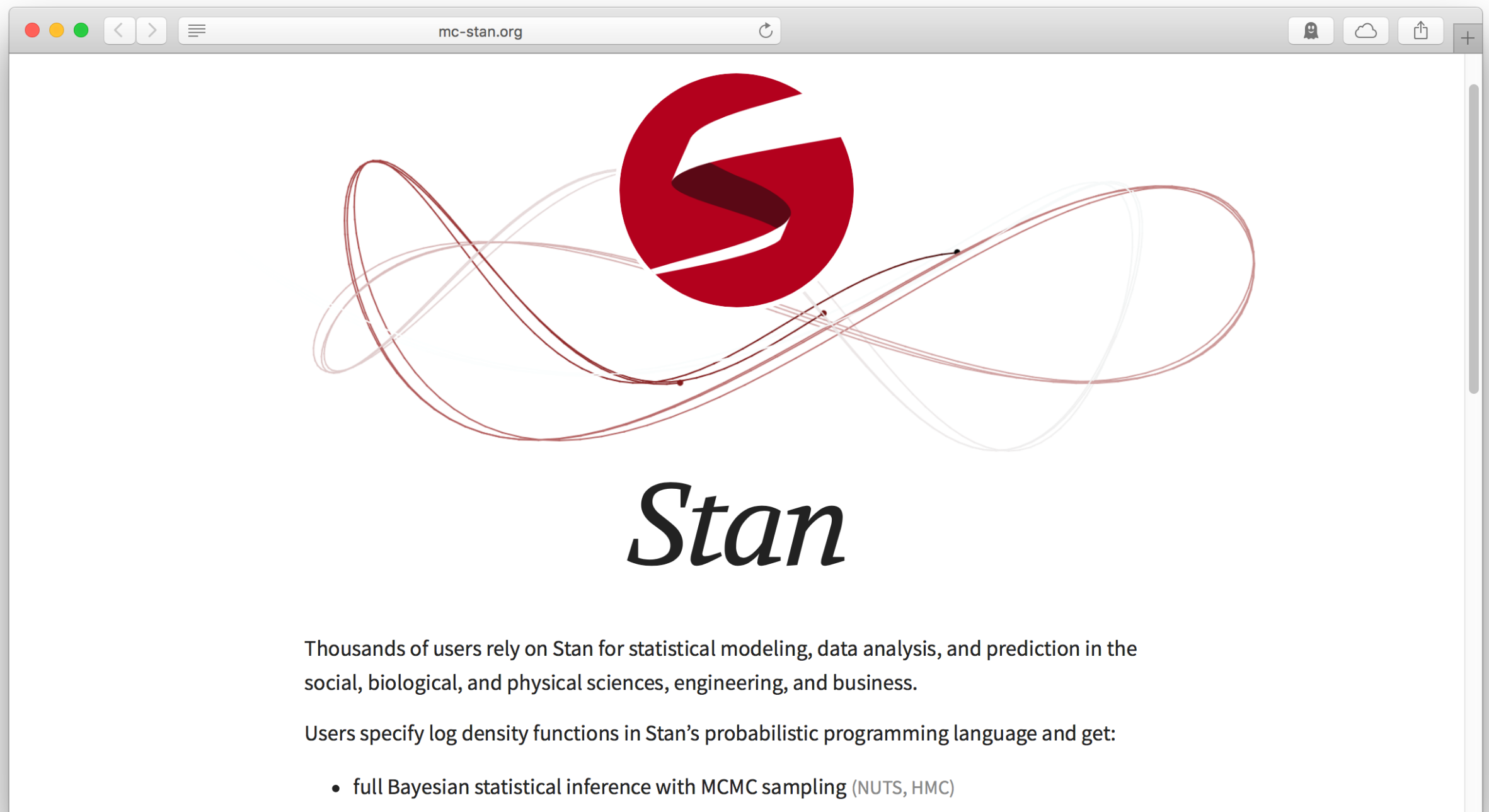
Count of Country–Years by Item





# Method

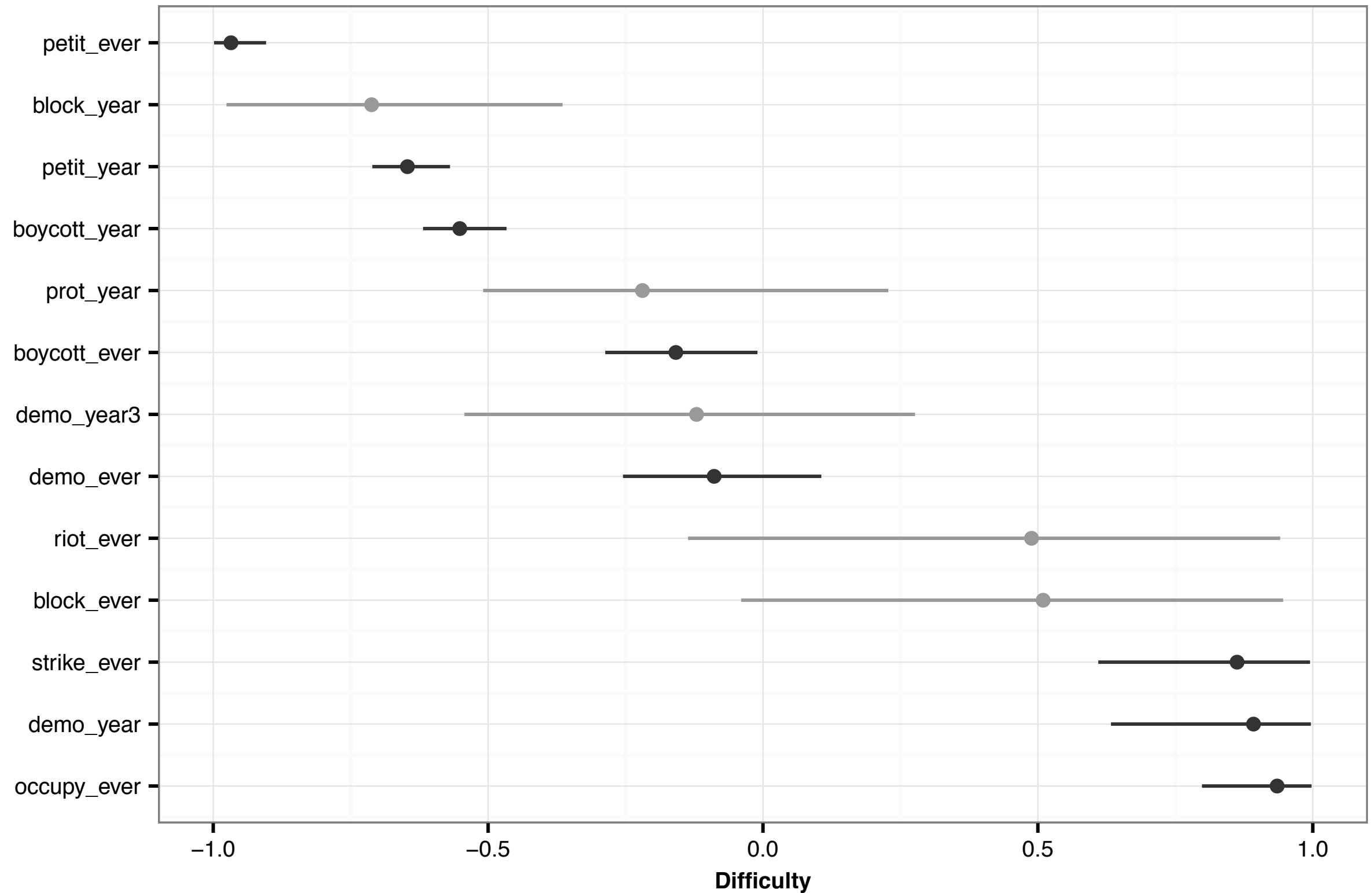
- Estimated using a Bayesian approach via Stan in R





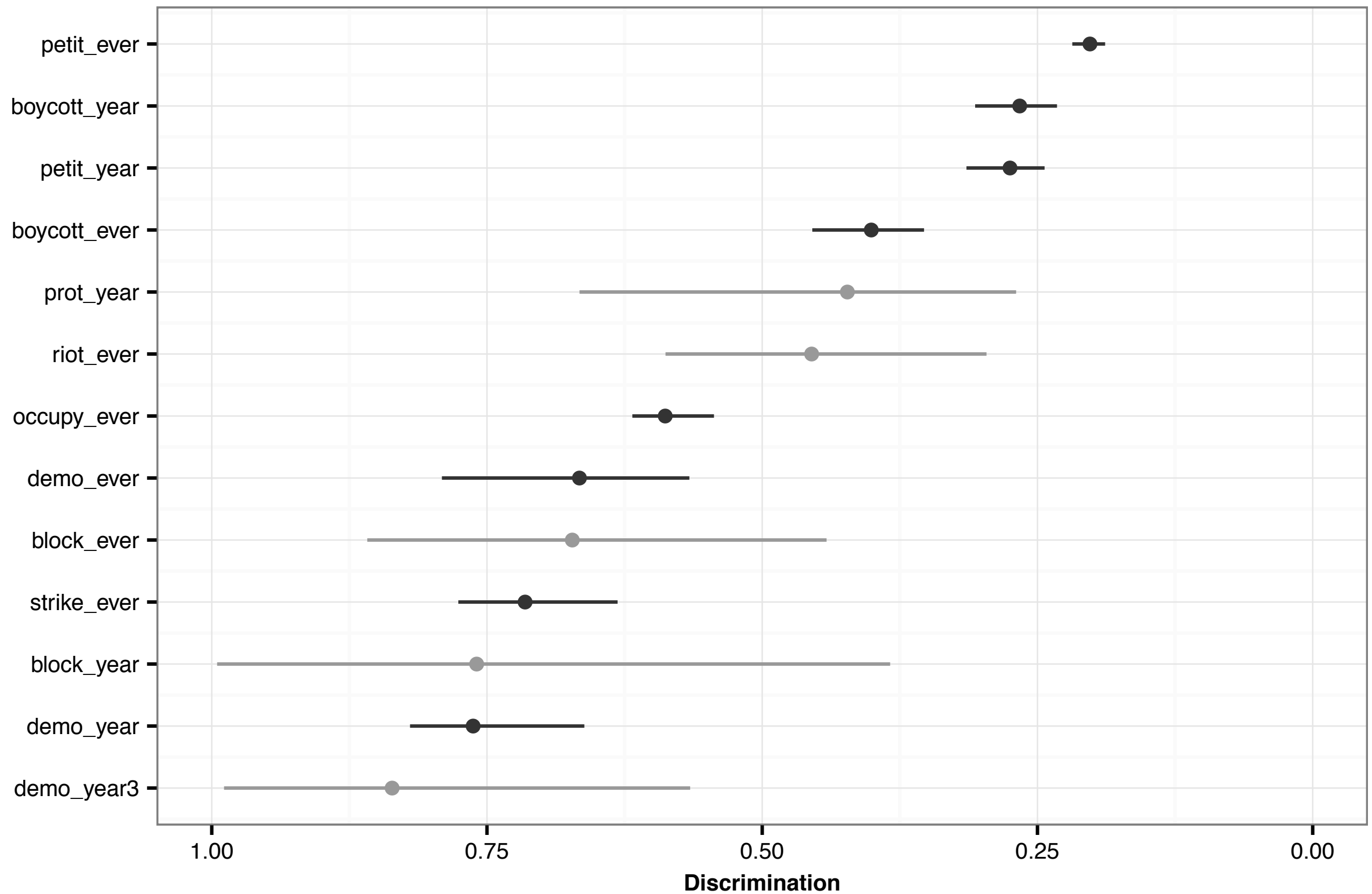
# Results

## Item Difficulty



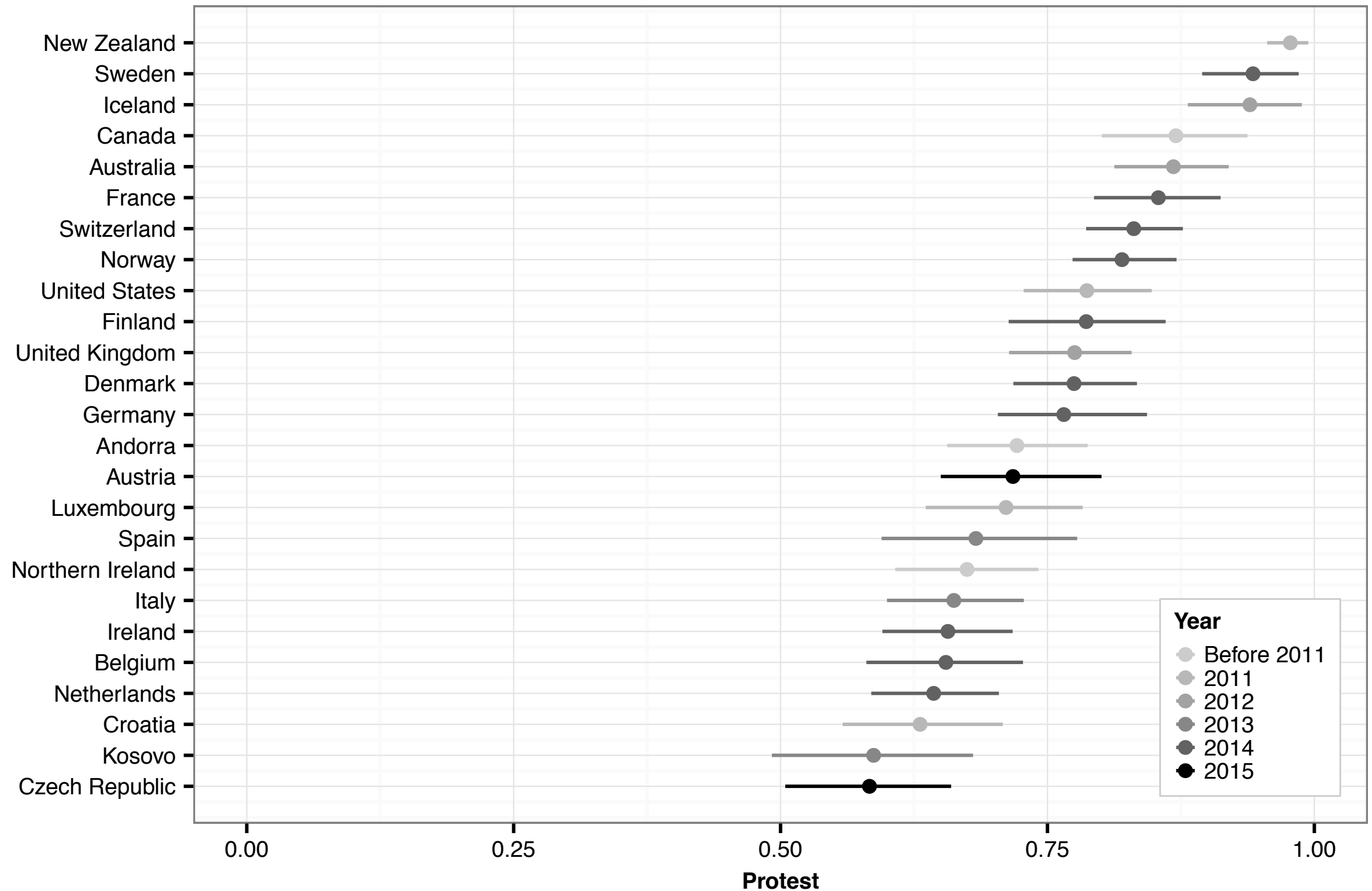
# Results

## Item Discrimination



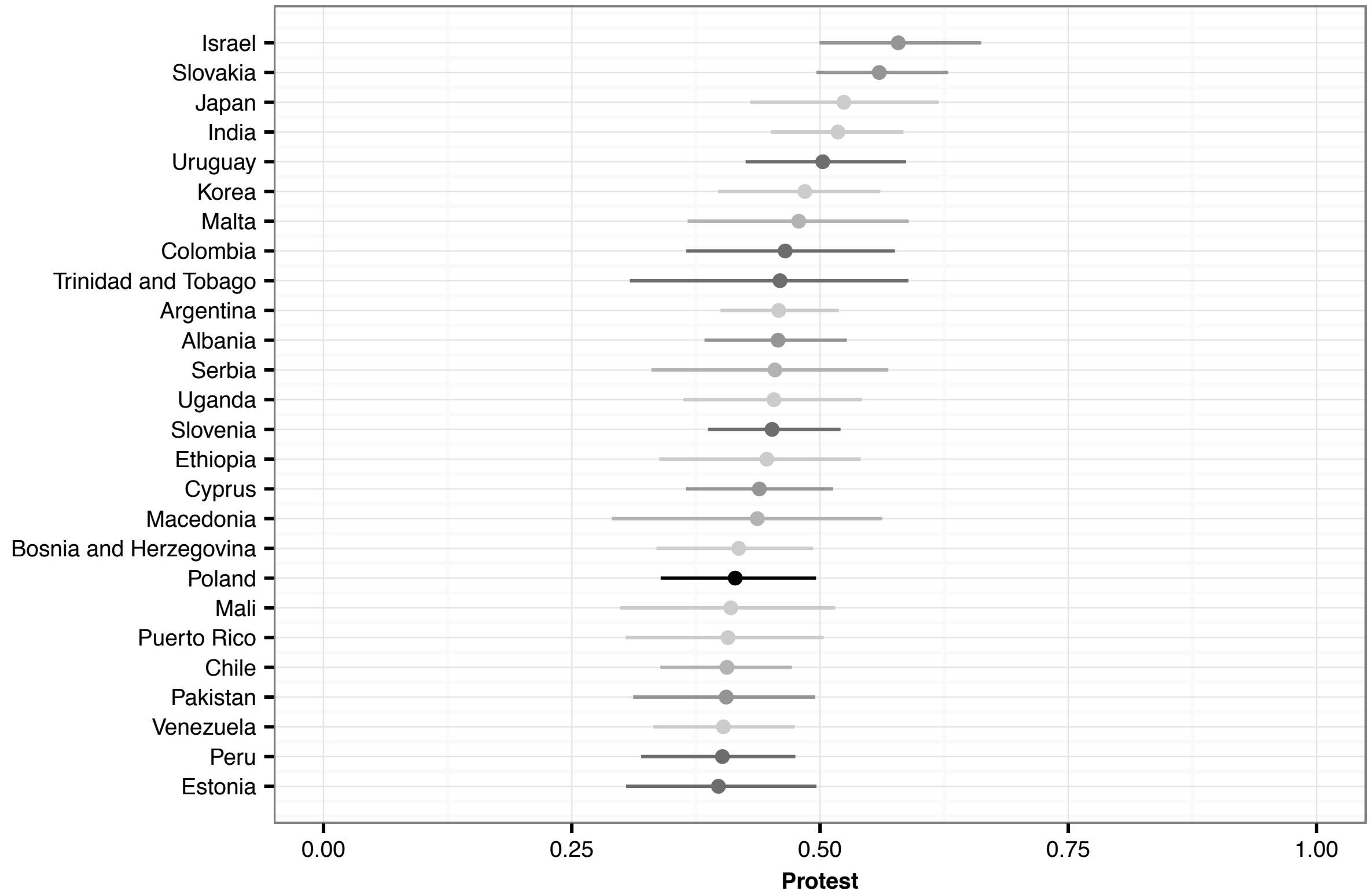
# Results

## A Cross-Section



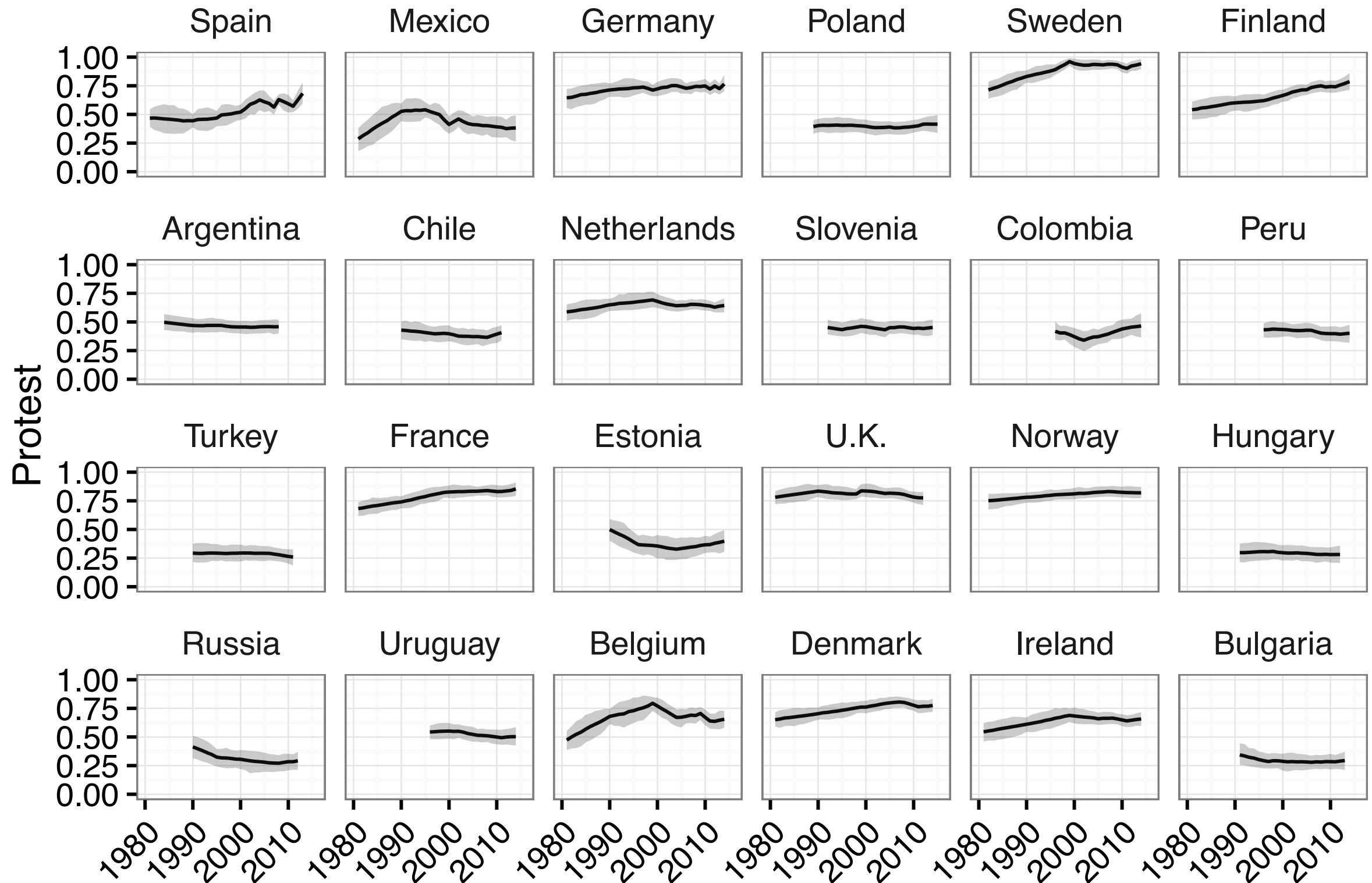
# Results

## A Cross-Section



# Results

## Time Series



# Conclusions

# Conclusions

- It's *possible* to estimate protest participation in this way

# Conclusions

- It's *possible* to estimate protest participation in this way
- Comparable TSCS data opens door to many interesting research questions ...



# Conclusions

- It's *possible* to estimate protest participation in this way
- But what exactly is being measured?
- Comparable TSCS data opens door to many interesting research questions ...

# Conclusions

- It's *possible* to estimate protest participation in this way
- Comparable TSCS data opens door to many interesting research questions ...
- But what exactly is being measured?
- Is timing appropriately modeled? (“ever” vs. “last 12 months”)

# Conclusions

- It's *possible* to estimate protest participation in this way
- Comparable TSCS data opens door to many interesting research questions ...
- But what exactly is being measured?
- Is timing appropriately modeled? (“ever” vs. “last 12 months”)