

Today's Agenda

What drives economic investment in US states?

- Introduce Dataset 1
- Begin univariate analyses

Justin Leinaweaver (Spring 2022)



Scientific models are:

- Neither true nor false
- Limited in their accuracy
- Partial representations
- Useful for only some uses
- A reflection of the interests of the designer

Dataset 1: The Motivating Problem

What drives economic investment in US states?

Why do some states attract greater investment by companies and individuals than others?

Dataset 1: The Motivating Problem

What drives economic investment in US states?

- ① Literature Review
- ② Exploratory Data Analysis

Dataset 1: Literature Review

What are the important causal mechanisms that explain business investment in states?

- 1 A focus group of educated, engaged and dynamic undergraduates interested in business
- 2 Girsch-Bock, Mary. (2021, Jan 19). Top 10 Best States to Start Your Small Business in 2021. *The Blueprint by The Motley Fool*.

Dataset 1: The Economies of the US States

| | A | B | C | D | E | F | G | H |
|----|-------------|--------|------|----------|--------------|--------------|----------------------|---------------------|
| 1 | State | abbrev | year | min_wage | gdp_millions | unemployment | population_thousands | rental_vacancy_rate |
| 2 | Alabama | AL | 2020 | | 224870.6 | 6 | 4921.532 | 13.6 |
| 3 | Alaska | AK | 2020 | 10.19 | 50246.7 | 7.9 | 731.158 | 7 |
| 4 | Arizona | AZ | 2020 | 12 | 372461 | 7.9 | 7421.401 | 5.5 |
| 5 | Arkansas | AR | 2020 | 10 | 129073.9 | 6.1 | 3030.522 | 8.4 |
| 6 | California | CA | 2020 | 12 | 3091871.5 | 10.2 | 39368.078 | 4 |
| 7 | Colorado | CO | 2020 | 12 | 390098.7 | 7.3 | 5807.719 | 4 |
| 8 | Connecticut | CT | 2020 | 12 | 280900.3 | 7.9 | 3557.006 | 5.6 |
| 9 | Delaware | DE | 2020 | 9.25 | 75512.5 | 7.9 | 986.809 | 6.1 |
| 10 | Florida | FL | 2020 | 8.56 | 1095888.2 | 7.9 | 21733.312 | 7.3 |
| 11 | Georgia | GA | 2020 | 7.25 | 619240 | 6.6 | 10710.017 | 7.1 |
| 12 | Hawaii | HI | 2020 | 10.1 | 89856.2 | 11.8 | 1407.006 | 7.5 |
| 13 | Idaho | ID | 2020 | 7.25 | 84032.2 | 5.4 | 1826.913 | 4.4 |
| 14 | Illinois | IL | 2020 | 10 | 863516.7 | 9.6 | 12587.53 | 7.9 |
| 15 | Indiana | IN | 2020 | 7.25 | 372636.7 | 7.2 | 6754.953 | 9.3 |
| 16 | Iowa | IA | 2020 | 7.25 | 192710.2 | 5.3 | 3163.561 | 8.9 |
| 17 | Kansas | KS | 2020 | 7.25 | 173298.3 | 5.9 | 2913.805 | 12.1 |
| 18 | Kentucky | KY | 2020 | 7.25 | 210024.2 | 6.6 | 4477.251 | 6.1 |

"Three Rules of Tidy Data"

| country | year | cases | population |
|-------------|------|--------|------------|
| Afghanistan | 1999 | 745 | 159807071 |
| Afghanistan | 2000 | 2666 | 20595360 |
| Brazil | 1999 | 37737 | 172006362 |
| Brazil | 2000 | 80488 | 174604898 |
| China | 1999 | 212258 | 127291272 |
| China | 2000 | 216766 | 128042583 |

variables

| country | year | cases | population |
|-------------|------|--------|------------|
| Afghanistan | 1999 | 745 | 159807071 |
| Afghanistan | 2000 | 2666 | 20595360 |
| Brazil | 1999 | 37737 | 172006362 |
| Brazil | 2000 | 80488 | 174604898 |
| China | 1999 | 212258 | 127291272 |
| China | 2000 | 216766 | 128042583 |

observations

| country | year | cases | population |
|-------------|------|--------|------------|
| Afghanistan | 1999 | 745 | 159807071 |
| Afghanistan | 2000 | 2666 | 20595360 |
| Brazil | 1999 | 37737 | 172006362 |
| Brazil | 2000 | 80488 | 174604898 |
| China | 1999 | 212258 | 127291272 |
| China | 2000 | 216766 | 128042583 |

values

Source: Wickham (2018) R for Data Science. O'Reilly.

Dataset 1: The Economies of the US States

| | A | B | C | D | E | F | G | H |
|----|-------------|--------|------|----------|--------------|--------------|----------------------|---------------------|
| 1 | State | abbrev | year | min_wage | gdp_millions | unemployment | population_thousands | rental_vacancy_rate |
| 2 | Alabama | AL | 2020 | | 224870.6 | 6 | 4921.532 | 13.6 |
| 3 | Alaska | AK | 2020 | 10.19 | 50246.7 | 7.9 | 731.158 | 7 |
| 4 | Arizona | AZ | 2020 | 12 | 372461 | 7.9 | 7421.401 | 5.5 |
| 5 | Arkansas | AR | 2020 | 10 | 129073.9 | 6.1 | 3030.522 | 8.4 |
| 6 | California | CA | 2020 | 12 | 3091871.5 | 10.2 | 39368.078 | 4 |
| 7 | Colorado | CO | 2020 | 12 | 390098.7 | 7.3 | 5807.719 | 4 |
| 8 | Connecticut | CT | 2020 | 12 | 280900.3 | 7.9 | 3557.006 | 5.6 |
| 9 | Delaware | DE | 2020 | 9.25 | 75512.5 | 7.9 | 986.809 | 6.1 |
| 10 | Florida | FL | 2020 | 8.56 | 1095888.2 | 7.9 | 21733.312 | 7.3 |
| 11 | Georgia | GA | 2020 | 7.25 | 619240 | 6.6 | 10710.017 | 7.1 |
| 12 | Hawaii | HI | 2020 | 10.1 | 89856.2 | 11.8 | 1407.006 | 7.5 |
| 13 | Idaho | ID | 2020 | 7.25 | 84032.2 | 5.4 | 1826.913 | 4.4 |
| 14 | Illinois | IL | 2020 | 10 | 863516.7 | 9.6 | 12587.53 | 7.9 |
| 15 | Indiana | IN | 2020 | 7.25 | 372636.7 | 7.2 | 6754.953 | 9.3 |
| 16 | Iowa | IA | 2020 | 7.25 | 192710.2 | 5.3 | 3163.561 | 8.9 |
| 17 | Kansas | KS | 2020 | 7.25 | 173298.3 | 5.9 | 2913.805 | 12.1 |
| 18 | Kentucky | KY | 2020 | 7.25 | 210024.2 | 6.6 | 4477.251 | 6.1 |

Dataset 1: The Motivating Problem

What drives economic investment in US states?

Why do some states attract greater investment by companies and individuals than others?

Dataset 1: The Economies of the US States

| | A | B | C | D | E | F |
|----|-------------|--------|------|--------------------------|--------------------------|----------------------|
| 1 | State | abbrev | year | gdp _millions | gdp _category | gdp _rate |
| 2 | Alabama | AL | 2020 | 224870.6 | Under \$1 trillion | -0.0143 |
| 3 | Alaska | AK | 2020 | 50246.7 | Under \$100 billion | -0.0761 |
| 4 | Arizona | AZ | 2020 | 372461 | Under \$1 trillion | 0.0063 |
| 5 | Arkansas | AR | 2020 | 129073.9 | Under \$1 trillion | -0.0144 |
| 6 | California | CA | 2020 | 3091871.5 | Above \$1 trillion | -0.0131 |
| 7 | Colorado | CO | 2020 | 390098.7 | Under \$1 trillion | -0.0073 |
| 8 | Connecticut | CT | 2020 | 280900.3 | Under \$1 trillion | -0.0240 |
| 9 | Delaware | DE | 2020 | 75512.5 | Under \$100 billion | -0.0204 |
| 10 | Florida | FL | 2020 | 1095888.2 | Above \$1 trillion | -0.0096 |
| 11 | Georgia | GA | 2020 | 619240 | Under \$1 trillion | -0.0103 |
| 12 | Hawaii | HI | 2020 | 89856.2 | Under \$100 billion | -0.0615 |
| 13 | Idaho | ID | 2020 | 84032.2 | Under \$100 billion | 0.0044 |
| 14 | Illinois | IL | 2020 | 863516.7 | Under \$1 trillion | -0.0249 |
| 15 | Indiana | IN | 2020 | 372636.7 | Under \$1 trillion | -0.0186 |
| 16 | Iowa | IA | 2020 | 192710.2 | Under \$1 trillion | -0.0100 |
| 17 | Kansas | KS | 2020 | 173298.3 | Under \$1 trillion | -0.0181 |

Dataset 1: The Economies of the US States

| | A | B | C | D | E | F | G | H |
|----|-------------|--------|------|----------|--------------|--------------|----------------------|---------------------|
| 1 | State | abbrev | year | min_wage | gdp_millions | unemployment | population_thousands | rental_vacancy_rate |
| 2 | Alabama | AL | 2020 | | 224870.6 | 6 | 4921.532 | 13.6 |
| 3 | Alaska | AK | 2020 | 10.19 | 50246.7 | 7.9 | 731.158 | 7 |
| 4 | Arizona | AZ | 2020 | 12 | 372461 | 7.9 | 7421.401 | 5.5 |
| 5 | Arkansas | AR | 2020 | 10 | 129073.9 | 6.1 | 3030.522 | 8.4 |
| 6 | California | CA | 2020 | 12 | 3091871.5 | 10.2 | 39368.078 | 4 |
| 7 | Colorado | CO | 2020 | 12 | 390098.7 | 7.3 | 5807.719 | 4 |
| 8 | Connecticut | CT | 2020 | 12 | 280900.3 | 7.9 | 3557.006 | 5.6 |
| 9 | Delaware | DE | 2020 | 9.25 | 75512.5 | 7.9 | 986.809 | 6.1 |
| 10 | Florida | FL | 2020 | 8.56 | 1095888.2 | 7.9 | 21733.312 | 7.3 |
| 11 | Georgia | GA | 2020 | 7.25 | 619240 | 6.6 | 10710.017 | 7.1 |
| 12 | Hawaii | HI | 2020 | 10.1 | 89856.2 | 11.8 | 1407.006 | 7.5 |
| 13 | Idaho | ID | 2020 | 7.25 | 84032.2 | 5.4 | 1826.913 | 4.4 |
| 14 | Illinois | IL | 2020 | 10 | 863516.7 | 9.6 | 12587.53 | 7.9 |
| 15 | Indiana | IN | 2020 | 7.25 | 372636.7 | 7.2 | 6754.953 | 9.3 |
| 16 | Iowa | IA | 2020 | 7.25 | 192710.2 | 5.3 | 3163.561 | 8.9 |
| 17 | Kansas | KS | 2020 | 7.25 | 173298.3 | 5.9 | 2913.805 | 12.1 |
| 18 | Kentucky | KY | 2020 | 7.25 | 210024.2 | 6.6 | 4477.251 | 6.1 |

| | | | | | | | | | | | | | | | | | | | | | | | |
|-------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| [1] | 237 | 222 | 186 | 115 | 200 | 232 | 164 | 131 | 125 | 145 | 217 | 165 | 164 | 181 | 175 | 233 | 243 | 233 | 138 | 146 | 147 | 211 | 133 |
| [24] | 185 | 164 | 127 | 193 | 189 | 178 | 249 | 181 | 231 | 162 | 108 | 186 | 243 | 193 | 106 | 124 | 105 | 114 | 237 | 111 | 163 | 196 | 174 |
| [47] | 182 | 240 | 244 | 166 | 164 | 190 | 236 | 204 | 122 | 121 | 135 | 190 | 244 | 210 | 201 | 246 | 121 | 127 | 229 | 154 | 109 | 113 | 169 |
| [70] | 241 | 191 | 121 | 171 | 105 | 205 | 159 | 192 | 209 | 167 | 166 | 169 | 219 | 146 | 218 | 104 | 220 | 137 | 132 | 218 | 160 | 229 | 208 |
| [93] | 156 | 244 | 160 | 112 | 129 | 136 | 135 | 234 | 214 | 176 | 137 | 228 | 159 | 245 | 170 | 215 | 195 | 118 | 110 | 228 | 133 | 123 | 159 |
| [116] | 235 | 166 | 221 | 224 | 141 | 123 | 200 | 209 | 142 | 235 | 212 | 246 | 153 | 217 | 118 | 232 | 104 | 111 | 213 | 229 | 149 | 243 | 139 |
| [139] | 186 | 159 | 150 | 127 | 110 | 191 | 108 | 248 | 184 | 244 | 145 | 190 | 122 | 222 | 153 | 180 | 242 | 122 | 135 | 142 | 228 | 149 | 131 |
| [162] | 176 | 229 | 108 | 223 | 119 | 132 | 197 | 242 | 154 | 201 | 103 | 149 | 197 | 106 | 111 | 185 | 163 | 120 | 218 | 166 | 245 | 132 | 123 |
| [185] | 205 | 181 | 211 | 138 | 231 | 233 | 158 | 100 | 209 | 231 | 159 | 242 | 225 | 233 | 168 | 206 | 136 | 154 | 221 | 110 | 113 | 126 | 152 |
| [208] | 117 | 102 | 240 | 218 | 120 | 124 | 133 | 150 | 222 | 210 | 146 | 213 | 142 | 193 | 174 | 197 | 178 | 154 | 132 | 183 | 145 | 156 | 104 |
| [231] | 150 | 172 | 147 | 200 | 128 | 231 | 139 | 240 | 213 | 125 | 183 | 158 | 169 | 180 | 196 | 127 | 210 | 240 | 123 | 141 | 203 | 128 | 203 |
| [254] | 218 | 110 | 176 | 114 | 139 | 144 | 183 | 192 | 154 | 116 | 225 | 103 | 156 | 222 | 157 | 113 | 105 | 229 | 110 | 115 | 149 | 249 | 105 |
| [277] | 224 | 188 | 143 | 118 | 108 | 137 | 224 | 204 | 135 | 178 | 174 | 153 | 122 | 126 | 206 | 154 | 105 | 178 | 192 | 152 | 241 | 250 | 212 |
| [300] | 173 | 187 | 193 | 144 | 211 | 208 | 223 | 136 | 139 | 105 | 166 | 182 | 137 | 137 | 195 | 240 | 187 | 211 | 150 | 104 | 216 | 119 | 183 |
| [323] | 246 | 199 | 142 | 171 | 189 | 160 | 180 | 165 | 191 | 139 | 126 | 182 | 100 | 240 | 205 | 111 | 115 | 235 | 116 | 209 | 159 | 125 | 245 |
| [346] | 182 | 145 | 159 | 186 | 177 | 241 | 182 | 178 | 138 | 247 | 226 | 181 | 180 | 225 | 145 | 233 | 200 | 175 | 244 | 153 | 127 | 174 | 133 |
| [369] | 200 | 231 | 172 | 177 | 108 | 240 | 220 | 230 | 188 | 153 | 249 | 164 | 217 | 240 | 154 | 227 | 167 | 149 | 234 | 241 | 101 | 164 | 243 |
| [392] | 246 | 230 | 140 | 187 | 135 | 114 | 131 | 118 | 165 | 153 | 191 | 167 | 190 | 205 | 162 | 228 | 210 | 135 | 203 | 125 | 105 | 245 | 223 |
| [415] | 106 | 126 | 155 | 244 | 147 | 171 | 153 | 169 | 180 | 103 | 104 | 135 | 119 | 162 | 155 | 174 | 139 | 222 | 122 | 162 | 126 | 148 | 230 |
| [438] | 190 | 170 | 158 | 113 | 172 | 114 | 200 | 110 | 152 | 188 | 218 | 241 | 156 | 131 | 133 | 115 | 186 | 237 | 134 | 123 | 180 | 195 | 205 |
| [461] | 147 | 190 | 101 | 214 | 162 | 204 | 140 | 148 | 141 | 153 | 156 | 176 | 139 | 225 | 111 | 128 | 200 | 248 | 117 | 191 | 182 | 119 | 192 |
| [484] | 134 | 165 | 164 | 109 | 178 | 133 | 187 | 146 | 204 | 169 | 207 | 146 | 140 | 141 | 194 | 163 | 225 | 149 | 113 | 203 | 217 | 150 | 244 |
| [507] | 170 | 238 | 143 | 122 | 186 | 234 | 190 | 106 | 224 | 160 | 125 | 177 | 143 | 141 | 175 | 233 | 171 | 115 | 110 | 117 | 144 | 125 | 226 |

| | | | | | | | | | | | |
|-------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| [1] | 237 | 222 | 186 | 115 | 200 | 232 | 164 | 131 | 125 | 145 | 217 |
| [12] | 165 | 164 | 181 | 175 | 233 | 243 | 233 | 138 | 146 | 147 | 211 |
| [23] | 133 | 185 | 164 | 127 | 193 | 189 | 178 | 249 | 181 | 231 | 162 |
| [34] | 108 | 186 | 243 | 193 | 106 | 124 | 105 | 114 | 237 | 111 | 163 |
| [45] | 196 | 174 | 182 | 240 | 244 | 166 | 164 | 190 | 236 | 204 | 122 |
| [56] | 121 | 135 | 190 | 244 | 210 | 201 | 246 | 121 | 127 | 229 | 154 |
| [67] | 109 | 113 | 169 | 241 | 191 | 121 | 171 | 105 | 205 | 159 | 192 |
| [78] | 209 | 167 | 166 | 169 | 219 | 146 | 218 | 104 | 220 | 137 | 132 |
| [89] | 218 | 160 | 229 | 208 | 156 | 244 | 160 | 112 | 129 | 136 | 135 |
| [100] | 234 | 214 | 176 | 137 | 228 | 159 | 245 | 170 | 215 | 195 | 118 |
| [111] | 110 | 228 | 133 | 123 | 159 | 235 | 166 | 221 | 224 | 141 | 123 |
| [122] | 200 | 209 | 142 | 235 | 212 | 246 | 153 | 217 | 118 | 232 | 104 |
| [133] | 111 | 213 | 229 | 149 | 243 | 139 | 186 | 159 | 150 | 127 | 110 |
| [144] | 191 | 108 | 248 | 184 | 244 | 145 | 190 | 122 | 222 | 153 | 180 |
| [155] | 242 | 122 | 135 | 142 | 228 | 149 | 131 | 176 | 229 | 108 | 223 |
| [166] | 119 | 132 | 197 | 242 | 154 | 201 | 103 | 149 | 197 | 106 | 111 |
| [177] | 185 | 163 | 120 | 218 | 166 | 245 | 132 | 123 | 205 | 181 | 211 |
| [188] | 138 | 231 | 233 | 158 | 100 | 209 | 231 | 159 | 242 | 225 | 233 |
| [199] | 168 | 206 | 136 | 154 | 221 | 110 | 113 | 126 | 152 | 117 | 102 |
| [210] | 240 | 218 | 120 | 124 | 133 | 150 | 222 | 210 | 146 | 213 | 142 |
| [221] | 193 | 174 | 197 | 178 | 154 | 132 | 183 | 145 | 156 | 104 | 150 |
| [232] | 172 | 147 | 200 | 128 | 231 | 139 | 240 | 213 | 125 | 183 | 158 |
| [243] | 169 | 180 | 196 | 127 | 210 | 240 | 123 | 141 | 203 | 128 | 203 |
| [254] | 218 | 110 | 176 | 114 | 139 | 144 | 183 | 192 | 154 | 116 | 225 |

- The middle?
- The range?
- The variation?

Defining Statistics: Level 1

Statistics is a set of tools we use to summarize data

Summarize: “give a brief statement of the main points of (something)” (Oxford Dictionary).

Defining Statistics: Level 2

“The practice or science of collecting and analyzing numerical data in large quantities, **especially for the purpose of inferring proportions in a whole from those in a representative sample**”
(Oxford Dictionary).

Descriptive Statistics (Johnson 2012)

Measures of Central Tendency

- Mean
- Median

Deviations from Central Tendency

- Standard deviation

Measures of Variability

- Range
- IQR

Descriptive Statistics (Johnson 2012)

Measures of Central Tendency

- Mean
- Median

Deviations from Central Tendency

- Standard deviation

Measures of Variability

- Range = Maximum - Minimum
- IQR = 75th - 25th percentile

Descriptive Statistics in Excel: Using Functions

| | A | B | C | D | E | F |
|---|-------------|------|--------------|---|----------------|----------|
| 1 | state | year | gdp_millions | | | |
| 2 | Alabama | 2018 | 221735.5 | | Mean | |
| 3 | Alaska | 2018 | 54734.1 | | GDP (millions) | 406455.9 |
| 4 | Arizona | 2018 | 348297.1 | | | |
| 5 | Arkansas | 2018 | 128418.9 | | | |
| 6 | California | 2018 | 2997732.8 | | | |
| 7 | Colorado | 2018 | 371749.6 | | | |
| 8 | Connecticut | 2018 | 275726.9 | | | |
| 9 | Delaware | 2018 | 73481.3 | | | |

Descriptive Statistics in Excel: Using Functions

| | A | B | C | D | E | F | G |
|---|-------------|------|--------------|--------------|---|----------------|----------|
| 1 | state | year | gdp_millions | gdp_billions | | | |
| 2 | Alabama | 2018 | 221735.5 | 221.7355 | | Mean | |
| 3 | Alaska | 2018 | 54734.1 | 54.7341 | | GDP (millions) | 406455.9 |
| 4 | Arizona | 2018 | 348297.1 | 348.2971 | | GDP (billions) | 406.4559 |
| 5 | Arkansas | 2018 | 128418.9 | 128.4189 | | | |
| 6 | California | 2018 | 2997732.8 | 2997.7328 | | | |
| 7 | Colorado | 2018 | 371749.6 | 371.7496 | | | |
| 8 | Connecticut | 2018 | 275726.9 | 275.7269 | | | |
| 9 | Delaware | 2018 | 73481.3 | 73.4813 | | | |

For Thursday

Variables

- Minimum wage
- Unemployment
- Population
- Homeowner Rate
- Manufacturing

Descriptive Statistics

- Mean
- Median
- Standard deviation
- Minimum
- Maximum
- 25th Percentile
- 75th Percentile

For Thursday

Predictors to Analyze (5): Min wage, unemployment, population, homeowner rate and manufacturing

| | |
|--------------------|----------------------------|
| Mean | = AVERAGE |
| Median | = MEDIAN |
| Standard deviation | = STDEV.S |
| Minimum | = MIN |
| Maximum | = MAX |
| 25th Percentile | = QUARTILE.EXC (quart = 1) |
| 75th Percentile | = QUARTILE.EXC (quart = 3) |
