Optimizing Your Data Given the Needs of Your Organization

Hannah Poquette & Aaron Butler

Kentucky Department of Education


I ♥ DATA
STRATEGIC DATA PROJECT

Overview

- Discuss how to most effectively utilize the data that is available in your organization for a variety of performance management needs.
- Present a framework for approaching data use and analysis.
- Showcase an example from our work at KDE.

The New York Times
The Upshot

There are large gaps between white children and their black and Hispanic classmates. The gaps are largest in places with large economic disparities.

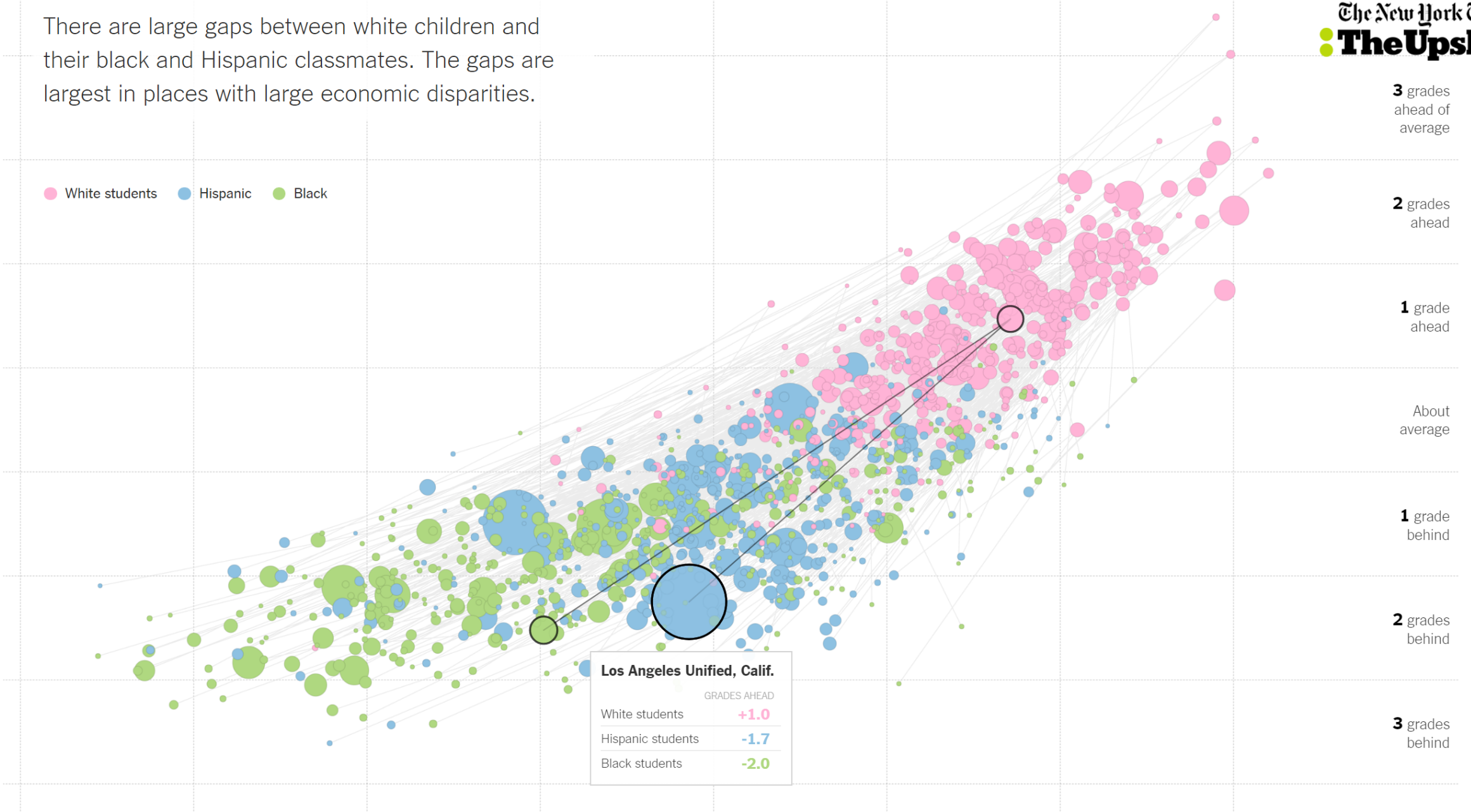


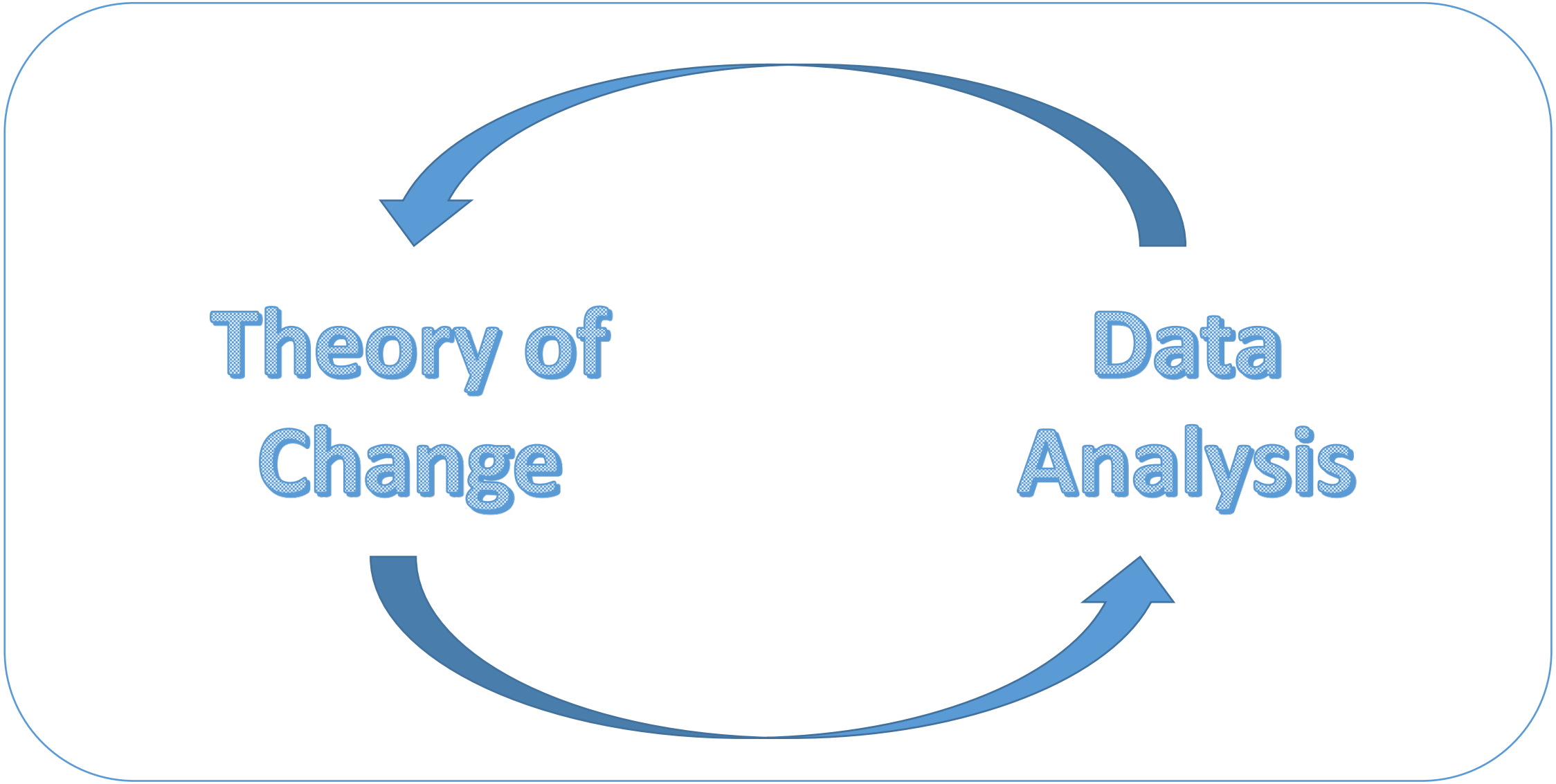
Chart shows districts with at least 100 white, 100 black and 100 Hispanic students per grade. Reliable estimates are not available for Asian-Americans.

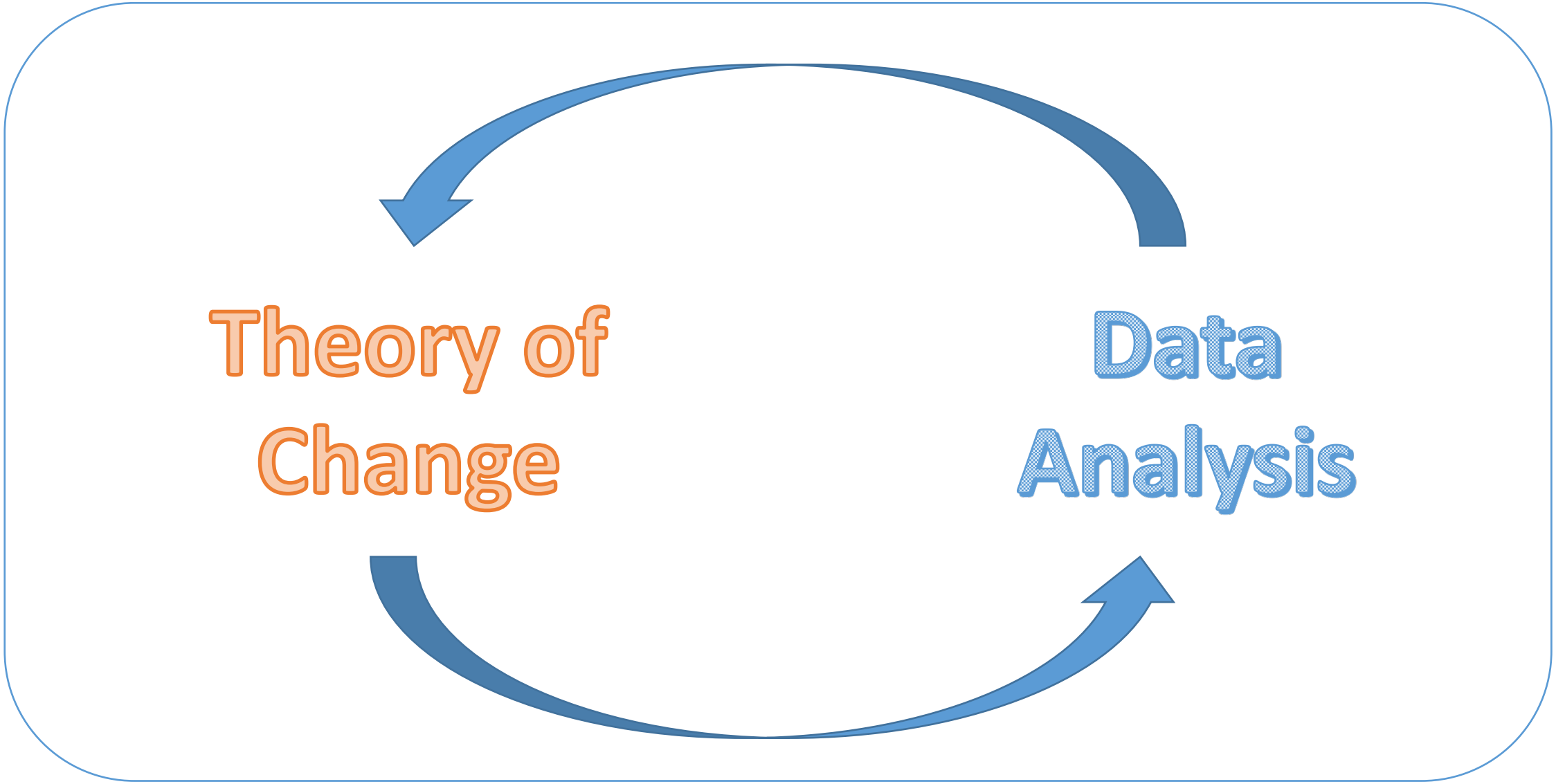
The goal is to turn data into information and information into insight.

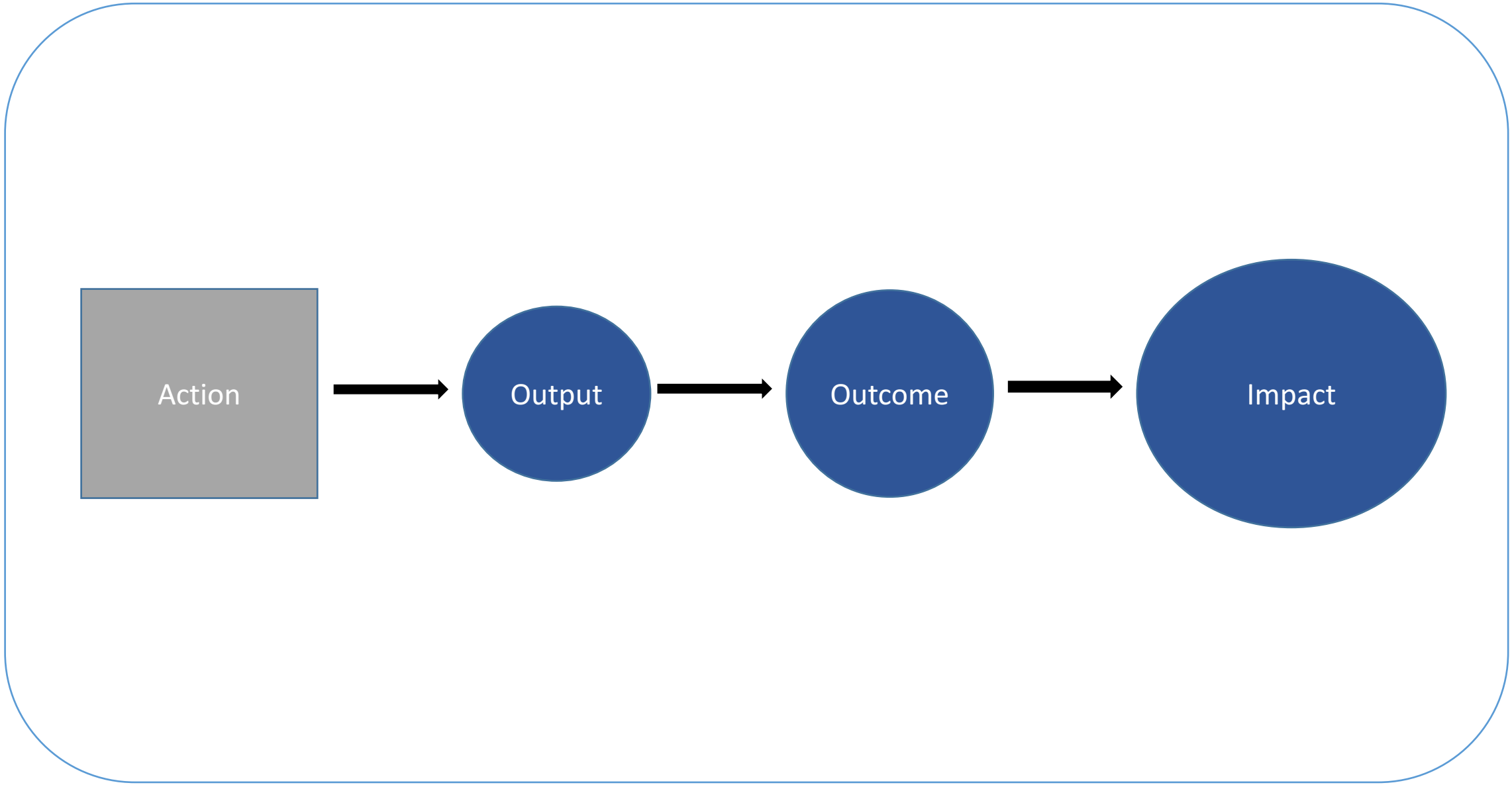
– Carly Fiorina –

How do we make data useful?

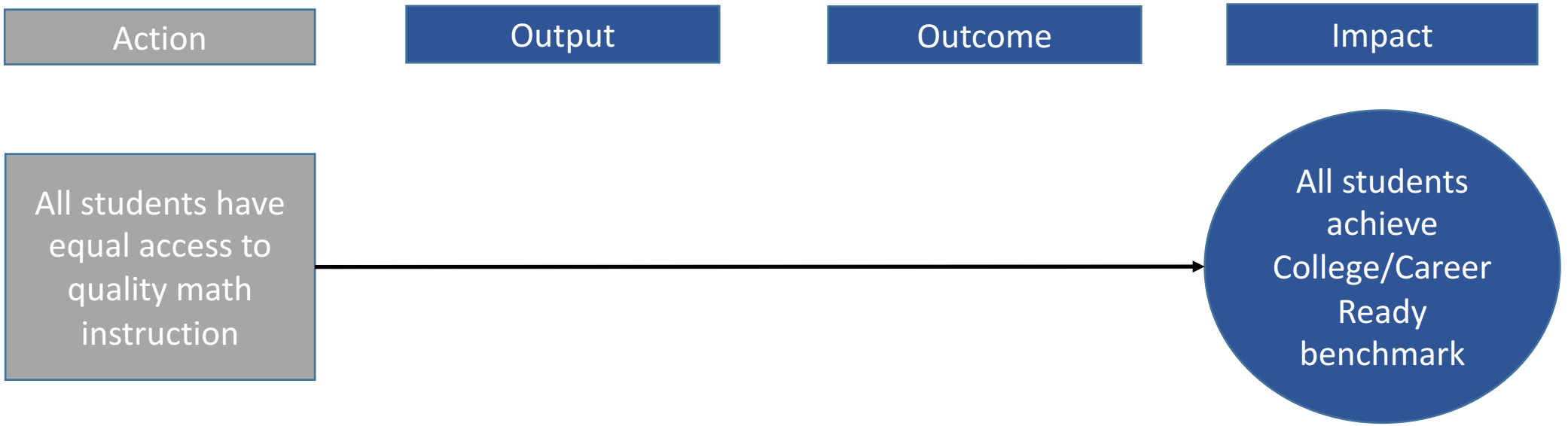
Do all students in Kentucky have an equitable opportunity to succeed?


Theory of Change
Data Analysis

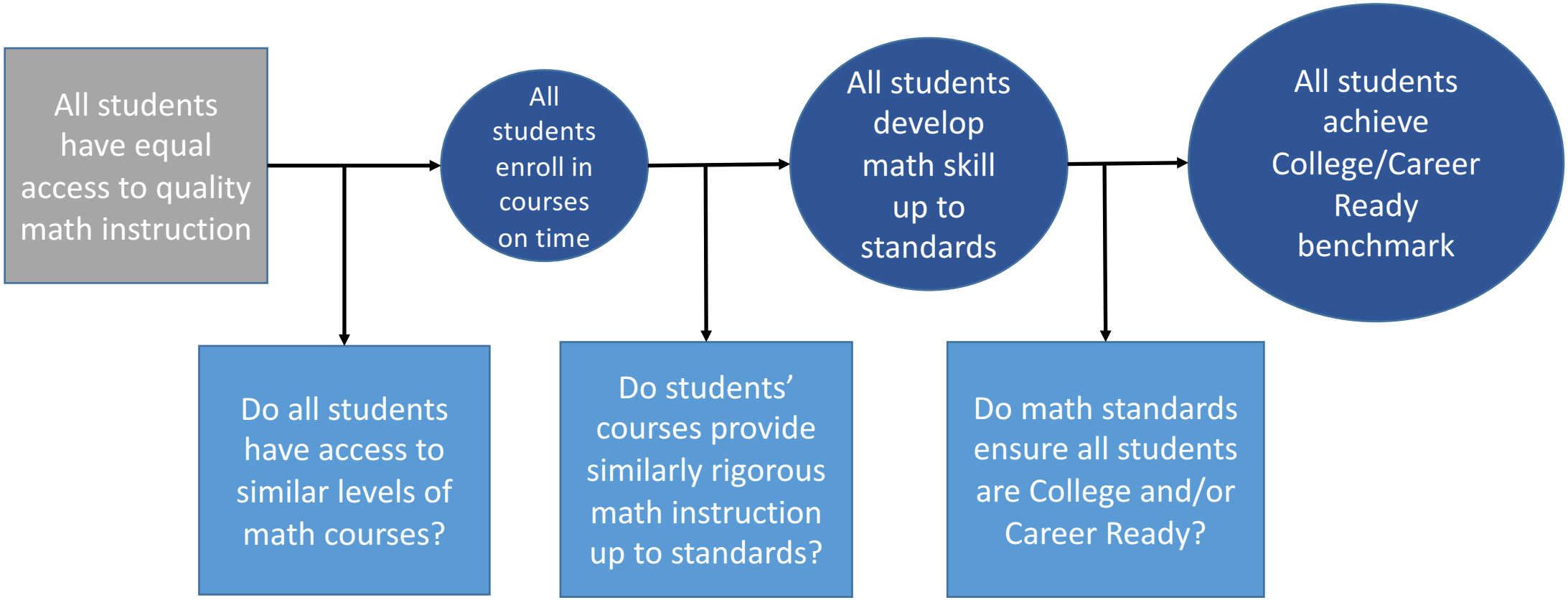

Theory of Change
Data Analysis

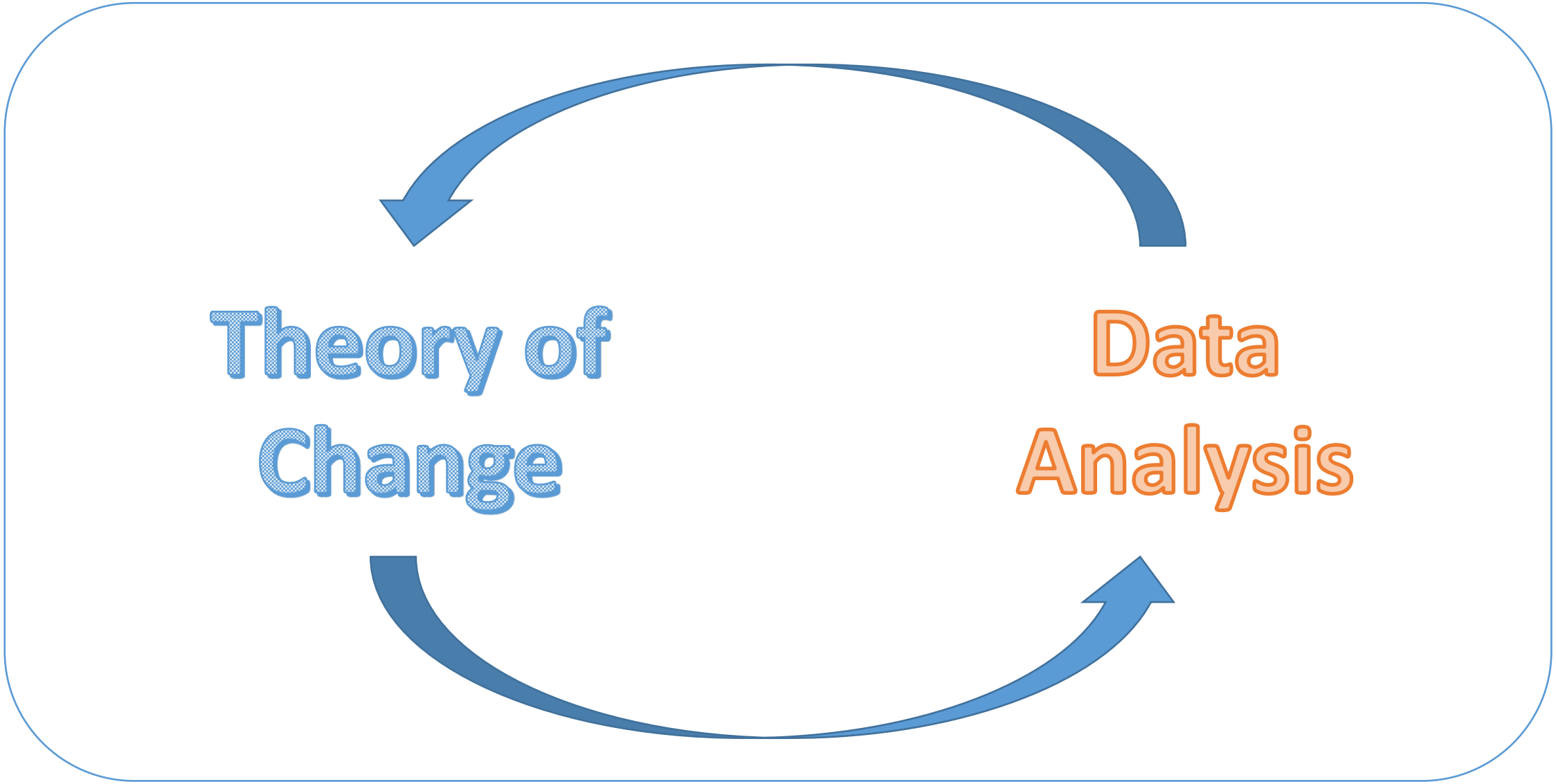

Action
Output
Outcome
Impact

Do all students in Kentucky have an equitable opportunity to succeed in math?



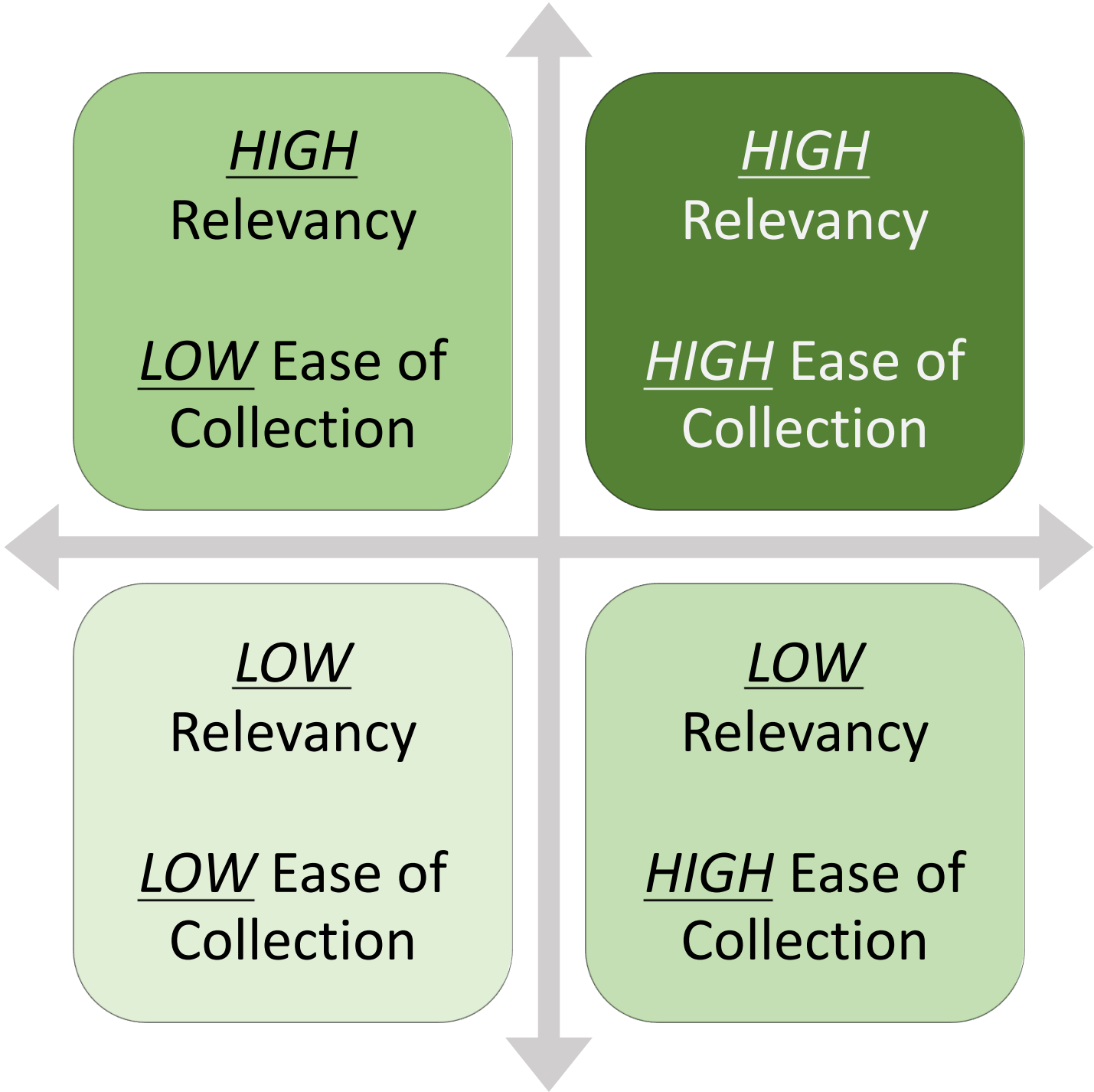
Do all students in Kentucky have an equitable opportunity to succeed in math?




Theory of Change
Data Analysis

Data Analysis


Gather
Process
Analyze
Report
Reflect


HIGH Relevancy
LOW Ease of Collection
HIGH Relevancy
HIGH Ease of Collection
LOW Relevancy
LOW Ease of Collection
LOW Relevancy
HIGH Ease of Collection

Research question: To what extent do student groups in Kentucky have similar levels of opportunity in their access to math courses?

- Math course enrollment
- Instructional materials
- Algebra 1 completion
- College/Career Readiness math benchmarks

What part of the TOC is most important data is available?

Refined question: Are student subgroups receiving similar levels of instructional quality in their math courses?

- Math GPA
- College/Career Readiness math benchmarks


Wallpaper Series


Gather
Process
Analyze
Report
Reflect

```r
# process data ------------------------------------------------------------

# clean column names
colnames(gr8_kprep)     <- col_lower(gr8_kprep)
colnames(act_data)      <- col_lower(act_data)
colnames(math_courses)  <- col_lower(math_courses)
colnames(enrollments)   <- col_lower(enrollments)
colnames(transcripts)   <- col_lower(transcripts)

# parse ms grades
ms_grade <- math_courses %>%
  mutate(ms_grade = grade_cleanr(final_grade)) %>%  # clean & convert to factor w/ function from function_help.R
  group_by(ssid) %>%
  summarize(avg_ms_grade = mean(as.numeric(ms_grade)),
            max_ms_grade = max(as.numeric(ms_grade))) # calculate avg/max grades

# parse hs grades
hs_grade <- transcripts %>%
  filter(course_year %in% c(2014, 2015, 2016)) %>% # limit to freshman, junior, and sophomore year
  mutate(hs_grade = grade_cleanr(score),   # process course grades w/ function_help.R
         credit_ratio = credit_earned / credit_attempted) %>% # create credit ratio var
  filter(!is.na(hs_grade)) %>% # filter out records w/o a grade
  filter(credit_ratio == 1 | credit_ratio == 0) %>% # filter out 14 records w/ partial credit oddities
  select(-credit_ratio) %>%
  # create hs math gpa
  group_by(ssid, course_year) %>%
  mutate(math_gpa_points = ((as.numeric(hs_grade) - 1) * credit_earned)) %>%
  summarise(math_gpa = sum(math_gpa_points) / sum(credit_attempted),
            gpa_points = sum(math_gpa_points),
            credit_attempt  = sum(credit_attempted))

# join ms grades + hs grades
df_grade <- ms_grade %>%
  left_join(hs_grade)

# join w/ gr8 math kprep data
```


Gather
Process
Analyze
Report
Reflect

```
> summary(fit_act19_hs2)

Call:
glm(formula = act_19_math ~ math_gpa + gender * race + iep +
    lep + frpl + as.factor(school_code), family = "binomial",
    data = df_out)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.7107  -0.7531  -0.2799   0.7659   3.6229

Coefficients:
                                                  Estimate Std. Error z value Pr(>|z|)
(Intercept)                                     -5.937e+00  8.323e-01  -7.134 9.78e-13 ***
math_gpa                                         1.198e+00  1.548e-02  77.377  < 2e-16 ***
genderF                                          9.950e-01  9.842e-01   1.011 0.312016
raceAsian                                        2.410e+00  8.030e-01   3.001 0.002691 **
raceBlack or African American                    8.098e-01  7.826e-01   1.035 0.300799
raceHispanic/Latino                              1.822e+00  7.853e-01   2.320 0.020350 *
raceNative Hawaiian or other Pacific Islander    2.207e+00  1.149e+00   1.920 0.054825 .
raceTwo or more races                            1.563e+00  7.900e-01   1.978 0.047925 *
raceWhite                                        1.863e+00  7.797e-01   2.389 0.016885 *
iepYes                                          -1.689e+00  6.291e-02 -26.850  < 2e-16 ***
lepYes                                          -2.298e+00  2.321e-01  -9.902  < 2e-16 ***
frplYes                                         -5.398e-01  2.720e-02 -19.845  < 2e-16 ***
as.factor(school_code)1012                      -1.186e+01  6.639e+02  -0.018 0.985749
as.factor(school_code)1014                       8.113e-01  3.952e-01   2.053 0.040085 *
as.factor(school_code)1016                      -1.351e+01  1.455e+03  -0.009 0.992595
as.factor(school_code)5015                       2.556e+00  1.333e+00   1.918 0.055086 .
as.factor(school_code)5020                      -1.208e+01  1.455e+03  -0.008 0.993376
as.factor(school_code)6010                       1.632e+01  1.019e+03   0.016 0.987228
as.factor(school_code)11020                      1.240e+00  3.430e-01   3.616 0.000300 ***
as.factor(school_code)11025                      9.453e-01  3.838e-01   2.463 0.013765 *
as.factor(school_code)11030                     -1.149e+01  1.455e+03  -0.008 0.993700
as.factor(school_code)12010                      1.227e+00  3.626e-01   3.384 0.000714 ***
as.factor(school_code)12011                     -1.095e+01  1.455e+03  -0.008 0.993996
as.factor(school_code)12050                      1.586e+00  4.159e-01   3.813 0.000137 ***
as.factor(school_code)13011                     -1.225e-01  7.262e-01  -0.169 0.866079
as.factor(school_code)15005                      1.115e+00  4.820e-01   2.312 0.020754 *
as.factor(school_code)15010                      1.351e+00  4.356e-01   3.101 0.001930 **
as.factor(school_code)16011                      1.122e+00  4.481e-01   2.505 0.012249 *
as.factor(school_code)17005                     -1.313e+01  9.380e+02  -0.014 0.988833
```

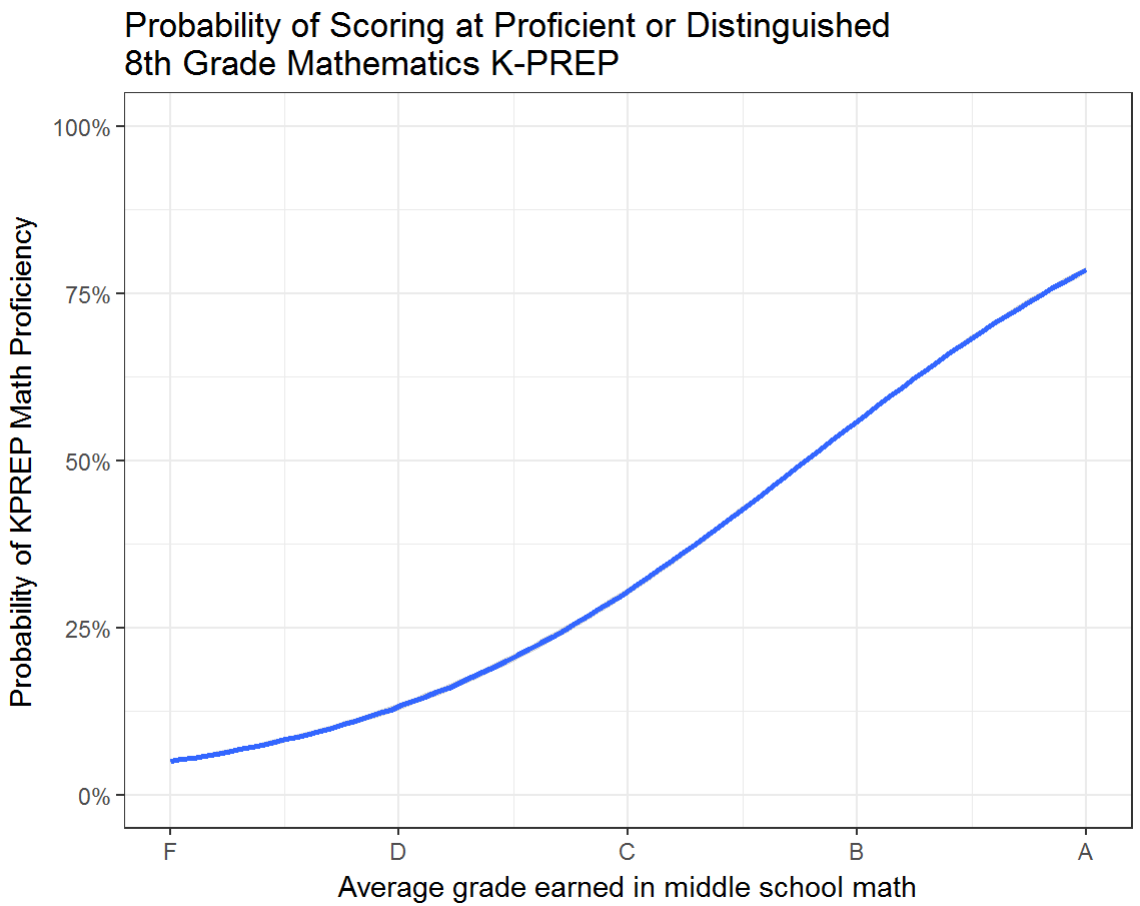
Table 8: Coefficients from the logistic regressions with interactions, ACT models

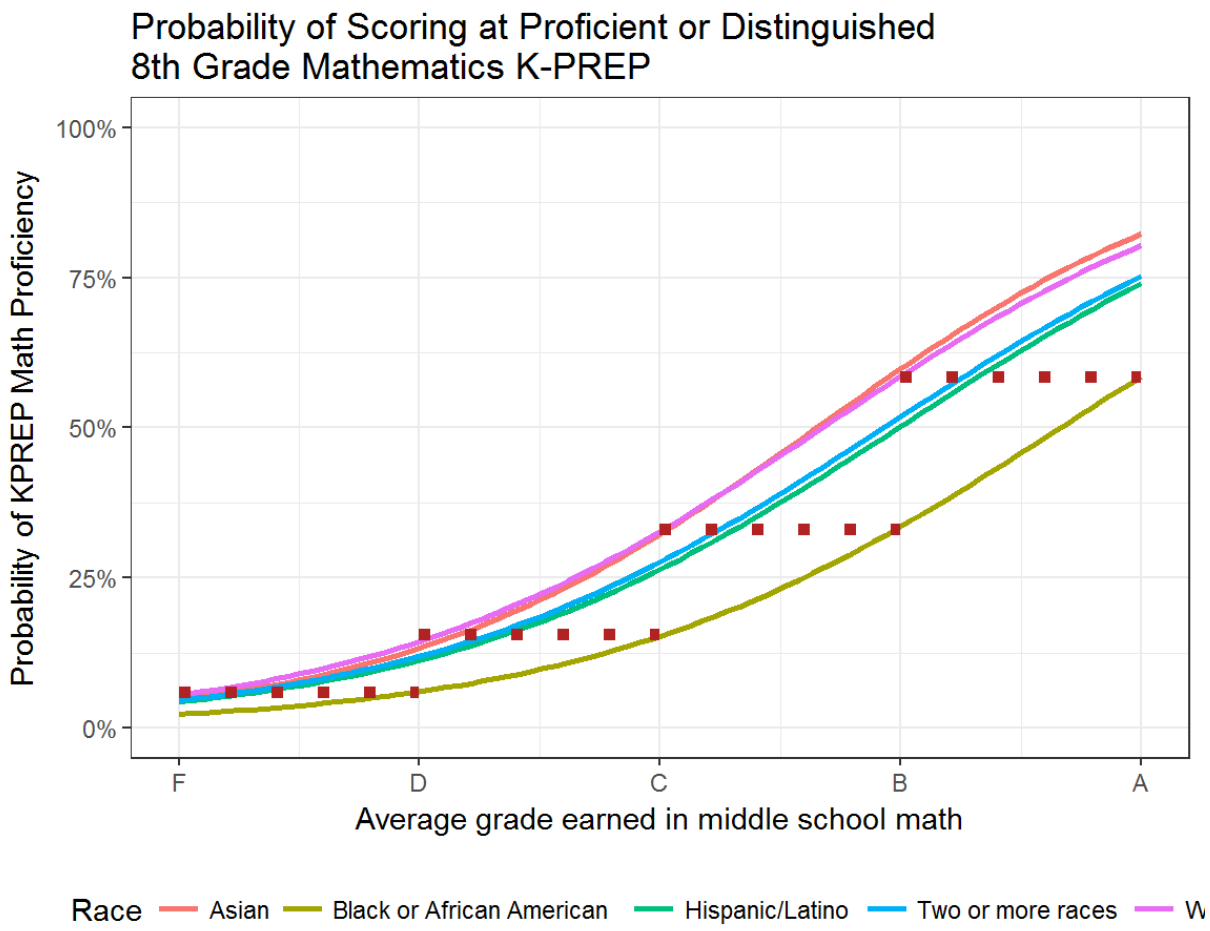
	ACT Mathematics Performance Level				
	(1)	(2)	(3)	(4)	(5)
High School Math Grade	1.078***	1.045***	1.036***	1.036***	1.198***
	(0.013)	(0.013)	(0.014)	(0.014)	(0.015)
Female		−0.383***	−0.468***	−0.468***	−0.547***
		(0.024)	(0.025)	(0.027)	(0.029)
American Indian		−0.772*	−0.792*	−1.451*	−1.863**
		(0.425)	(0.424)	(0.756)	(0.780)
Asian		0.496***	0.844***	0.991***	0.547***
		(0.105)	(0.120)	(0.182)	(0.195)
Black/African American		−0.738***	−0.738***	−0.829***	−1.053***
		(0.045)	(0.046)	(0.067)	(0.075)
Hispanic/Latino		−0.140**	0.004	0.161*	−0.041
		(0.063)	(0.067)	(0.093)	(0.099)
Native Hawaiian		−0.099	−0.026	0.666	0.344
		(0.500)	(0.532)	(0.769)	(0.845)
Two or more races		−0.048	−0.080	−0.103	−0.300**
		(0.080)	(0.081)	(0.123)	(0.131)
FRPL		−0.909***	−0.836***	−0.836***	−0.540***
		(0.024)	(0.025)	(0.025)	(0.027)
IEP			−1.622***	−1.620***	−1.689***
			(0.059)	(0.059)	(0.063)
LEP			−2.366***	−2.394***	−2.298***
			(0.219)	(0.221)	(0.232)
Constant	−3.411***	−2.605***	−2.460***	−2.460***	−4.074***
	(0.039)	(0.042)	(0.043)	(0.043)	(0.292)
Interactions	No	No	No	Yes	Yes
School Fixed Effects	No	No	No	No	Yes
Observations	40,648	40,648	40,648	40,648	40,648
Log Likelihood	−22,509.240	−21,320.390	−20,739.950	−20,732.780	−19,034.460
Akaike Inf. Crit.	45,022.480	42,660.770	41,503.900	41,501.560	39,508.920

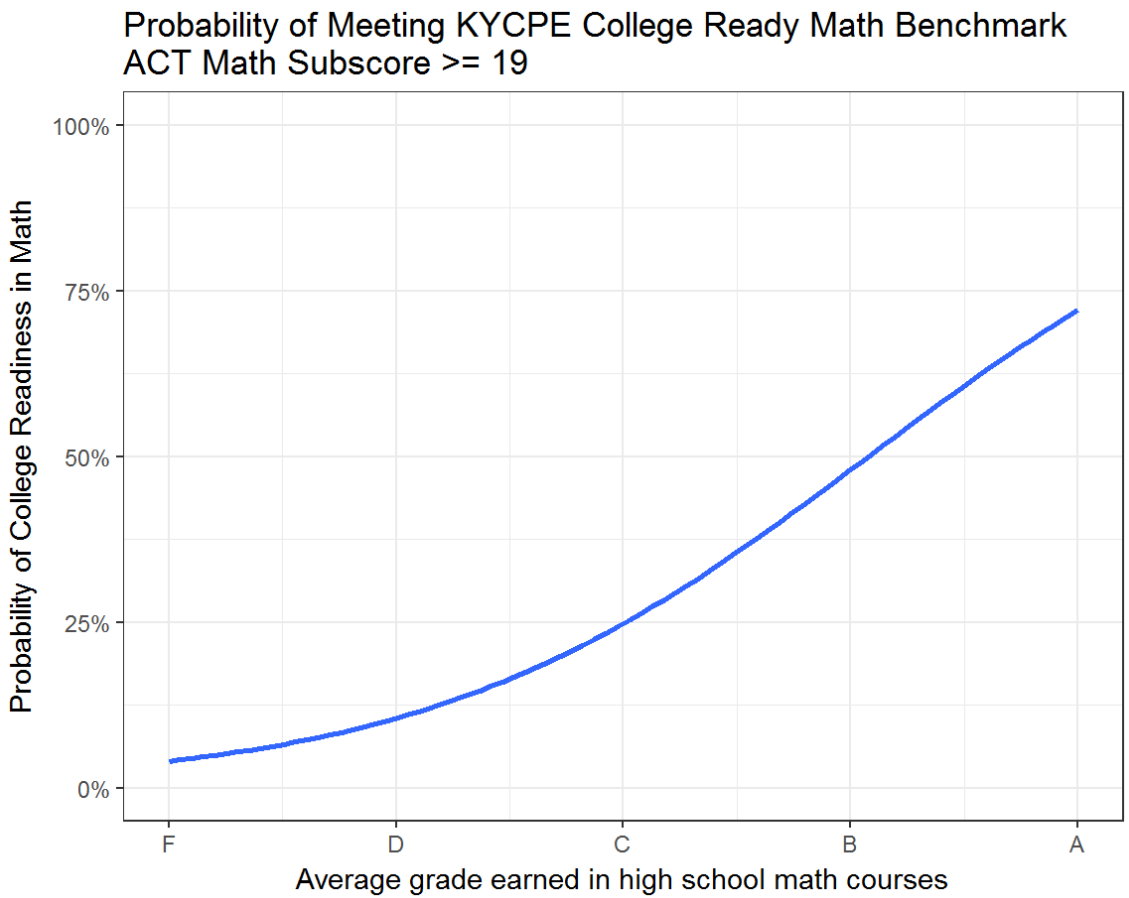
Note: *p<0.1; **p<0.05; ***p<0.01

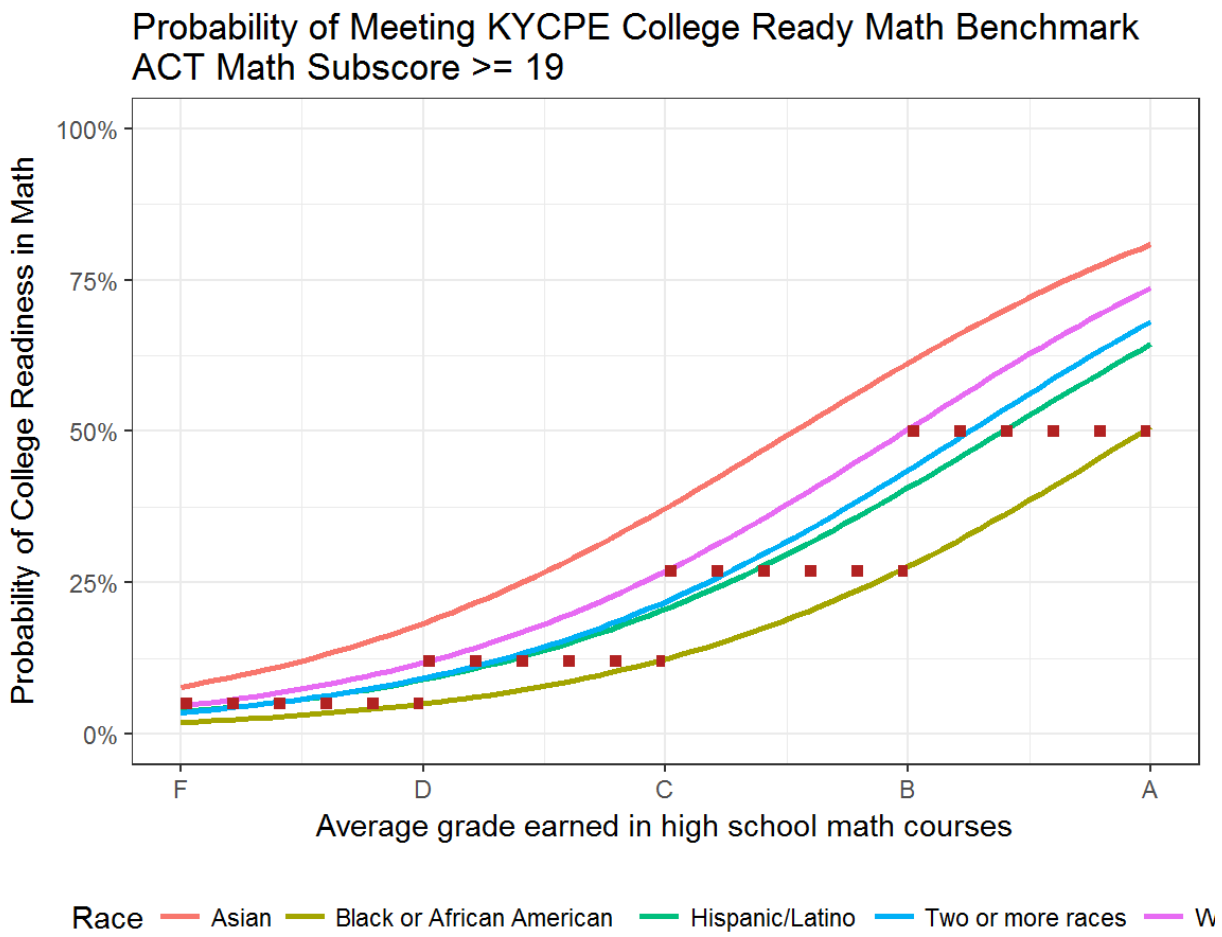
Reference group: male, white, no IEP, no LEP, no FRPL

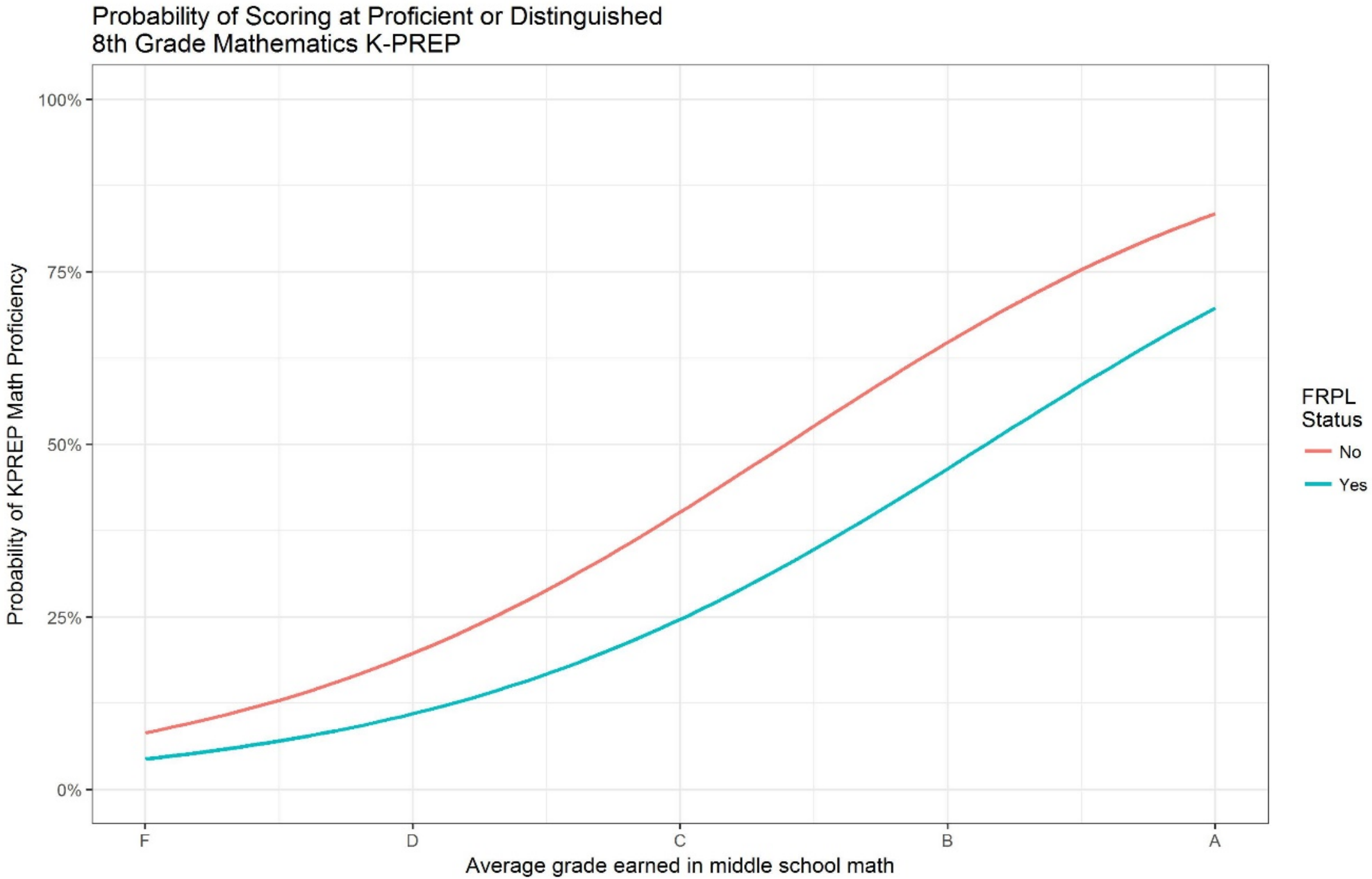

Gather
Process
Analyze
Report
Reflect


Probability of Scoring at Proficient or Distinguished
8th Grade Mathematics K-PREP
Probability of KPREP Math Proficiency
100%
75%
50%
25%
0%
F
D
C
B
A
Average grade earned in middle school math

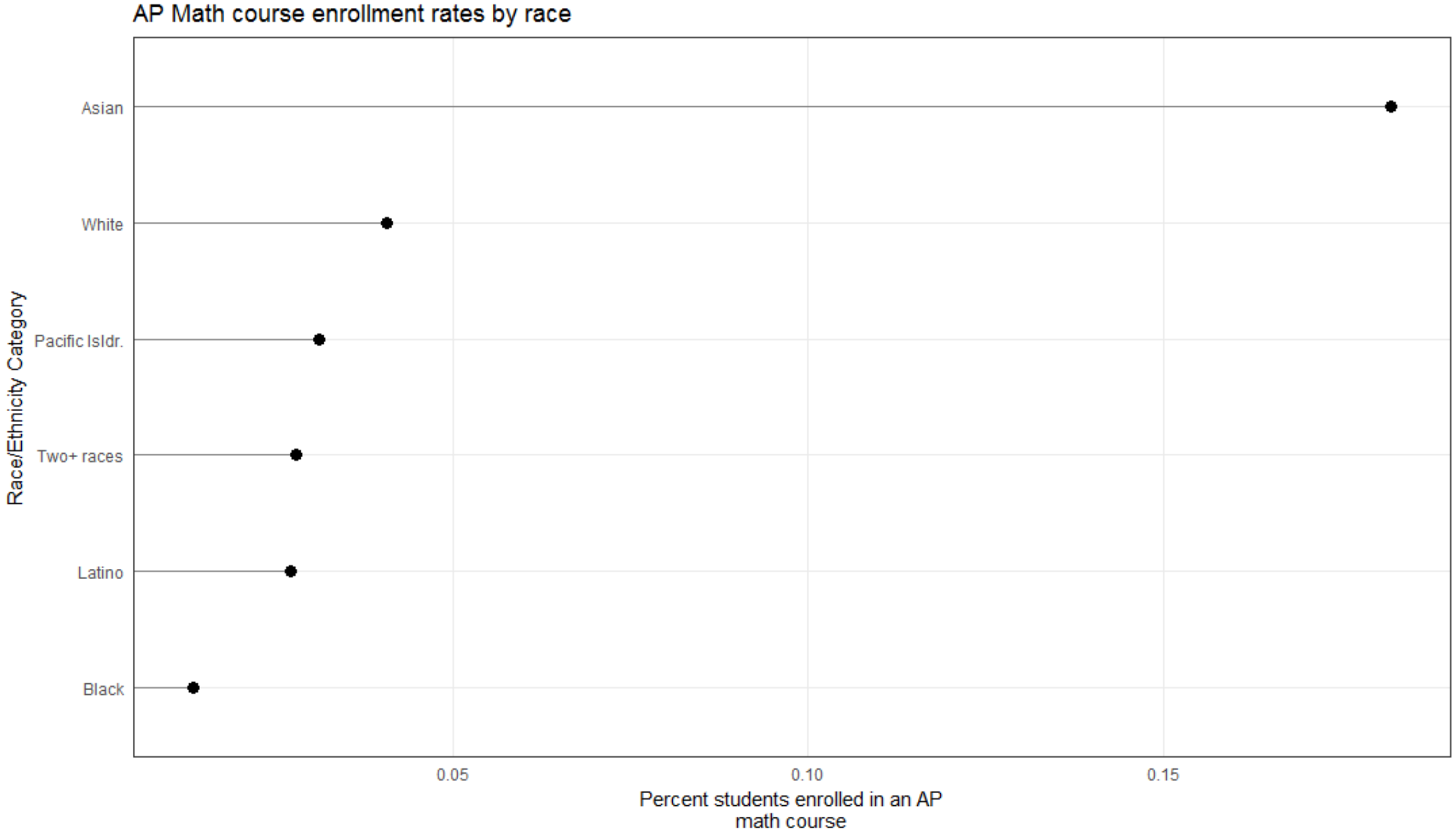

Probability of Scoring at Proficient or Distinguished
8th Grade Mathematics K-PREP
Probability of KPREP Math Proficiency
100%
75%
50%
25%
0%
F
D
C
B
A
Average grade earned in middle school math
Race
Asian
Black or African American
Hispanic/Latino
Two or more races
W


Probability of Meeting KYCPE College Ready Math Benchmark
ACT Math Subscore >= 19
Probability of College Readiness in Math
100%
75%
50%
25%
0%
F
D
C
B
A
Average grade earned in high school math courses


Probability of Meeting KYCPE College Ready Math Benchmark
ACT Math Subscore >= 19
Probability of College Readiness in Math
100%
75%
50%
25%
0%
F
D
C
B
A
Average grade earned in high school math courses
Race
Asian
Black or African American
Hispanic/Latino
Two or more races
W


Probability of Scoring at Proficient or Distinguished
8th Grade Mathematics K-PREP
100%
75%
50%
25%
0%
Probability of KPREP Math Proficiency
F
D
C
B
A
Average grade earned in middle school math
FRPL Status
No
Yes


Gather
Process
Analyze
Report
Reflect


AP Math course enrollment rates by race
Asian
White
Pacific Isldr.
Two+ races
Latino
Black
Race/Ethnicity Category
0.05
0.10
0.15
Percent students enrolled in an AP math course

Summary

- Identify your theory of change and underlying assumptions
- Assess the relevancy and availability of your desired data
- Determine what questions you can answer given the limitations of your data
- Document, document, document…
- Tell a story (with a clear narrative)
- Reflect on the implications of your results (i.e., the ‘so what’)

We answer emails…

Hannah Poquette
hannah.poquette@education.ky.gov

Aaron Butler
aaron.butler@education.ky.gov