

Equity Data: Using SAS® to Establish the Baseline

Smith, Kelly D., Central Piedmont Community College

ABSTRACT

Educational institutions across the United States are taking steps to achieve equity in access, progress, and outcomes for all students. Progress toward equity is measured against a starting baseline in order to effectively evaluate the efficacy and efficiency of institutional efforts. SAS® offers several options for exploring equity data to establish the baseline against which equity efforts are measured.

The presentation, appropriate for all levels of SAS users, will demonstrate the utility of SAS for analyzing and presenting equity data using Base SAS and SAS Enterprise Guide. The discussion will include PROC FREQ, PROC TABULATE, PROC REPORT, PROC SGPLOT, and SGPANEL. Code and visualization examples will be shared during the presentation.

INTRODUCTION

At Central Piedmont Community College, a working group (CP Economic Mobility Roundtable, 2019) defined education equity as “the institutional process of addressing and diminishing disparities in access, retention, persistence, completion, and success for groups underrepresented and underserved by educational institutions (due to the impact of historical, societal and cultural barriers).”

A wide variety of policy actors are working to establish educational equity in K12 and postsecondary institutions (Center for American Progress, 2020; Fink & Jenkins, 2020, Organization for Economic Co-operation and Development, 2018; Peters, 2020; Sargrad, et al., 2019). Covid-19 has only emphasized the current inequities within the American educational system (Simmons, 2020) as remote learning has become the norm during the pandemic.

An important first step in working toward educational equity is establishing the baseline. The baseline is determined by the current state of student enrollment and success at a particular education institution. In this paper, student data is modeled with the sashelp.baseball data set. Five variables from the data set were used to help create the visualizations: Name (character), Position (character), logSalary (numeric), nAtBat (numeric), and nHits (numeric). Four analysis groups were created from the Position variable; GPA data was modeled with logSalary, and a success variable, PctHts (numeric) was computed from nHits and nAtBat ($PctHts = (nHits/nAtBat) * 100$).

SAS® offers a variety of procedures for data analysis and visualization; each procedure can be further customized through the use of options. The choice of procedure depends on the question being answered, the type of variable being analyzed, and the audience for which the analysis is being completed.

COHORT COMPOSITION

Before student success is evaluated, it is important to know the composition of the student groups in the equity efforts. The PROC FREQ and PROC SGPLOT procedures offer a simple way to provide the essential information needed to start the analysis:

```
TITLE      "Analysis Groups for Equity";  
TITLE2    "PROC FREQ";
```

```

PROC FREQ DATA=BB02;
    TABLES Group;
RUN;

TITLE    "Analysis Groups for Equity";
TITLE2   "PROC SGPLOT";
PROC SGPLOT DATA=BB02;
    VBAR Group;
RUN;

```

Figure 1 displays the output created by the base SAS code. A similar visualization can be created with the PROC SQL procedure by using the “One-Way Frequencies” option under the Describe tab in Enterprise Guide.

Analysis Groups for Equity
PROC FREQ

Group	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Base Player	80	38.46	80	38.46
Catcher	30	14.42	110	52.88
Designated Hitter	11	5.29	121	58.17
Fielder	87	41.83	208	100.00

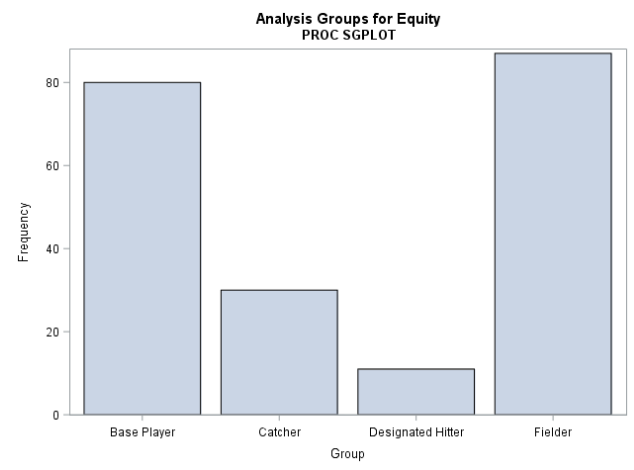


Figure 1. Cohort Composition Visualizations

COHORT CHARACTERISTICS - NUMERIC

A common student characteristic used in the analysis of equity is GPA (grade point average). The PROC MEANS procedure provides many options for univariate analysis for the entire student cohort or for subgroups. By using the PROC TABULATE procedure, information for both the entire cohort and subgroups can be combined into one table. The PROC SGPLOT procedure can create a histogram of student GPA and compare the resulting histogram to a normal distribution. PROC SGPLOT can also be used to create a box and whiskers plot:

```

TITLE    "Baseball Salary Distribution";
TITLE2   "PROC MEANS";
PROC MEANS DATA=BB02 NAXDEC=2;
    VAR logSal;
RUN;

TITLE    "Baseball Salary Distribution";
TITLE2   "PROC TABULATE";
PROC TABULATE DATA=BB02;
    CLASS Group;
    VAR logSal;
    TABLE Group (ALL), logSal="log Salary" * (N Min Max Mean STD)
RUN;

```

```

TITLE      "Baseball Salary Distribution";
TITLE2    "PROC SGPLOT";
PROC SGPLOT DATA=BB02;
    HISTOGRAM logSal;
    DENSITY logSal;
RUN;

```

```

TITLE      "Baseball Salary Distribution";
TITLE2    "PROC SGPLOT";
PROC SGPLOT DATA=BB02;
    VBOX logSal / CATEGORY Group;
RUN;

```

Figure 2 displays the output created by the base SAS code. Similar visualizations can be created in Enterprise Guide by using the “Summary Statistics Wizard” and “Summary Tables Wizard” options under the Describe tab.

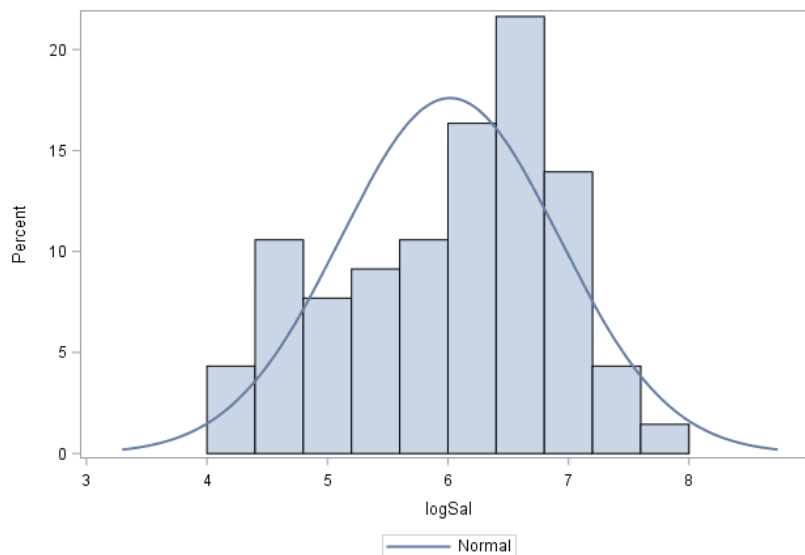
**Baseball Salary Distribution
PROC MEANS**

Analysis Variable : logSal				
N	Mean	Std Dev	Minimum	Maximum
208	6.02	0.91	4.21	7.81

**Baseball Salary Distribution
PROC TABULATE**

	log Salary				
	N	Min	Max	Mean	Std
Group					
Base Player	80	4.21	7.81	6.08	0.94
Catcher	30	4.32	7.56	5.90	0.92
Designated Hitter	11	4.98	7.00	6.18	0.58
Fielder	87	4.22	7.79	5.98	0.91
All	208	4.21	7.81	6.02	0.91

**Baseball Salary Distribution
PROC SGPLOT**



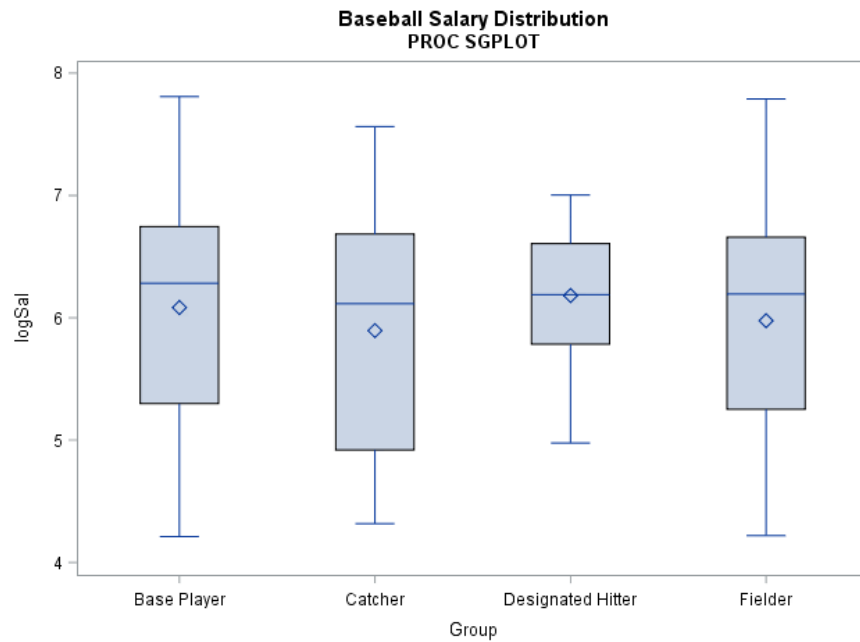


Figure 2. Analysis and Visualizations of a Numeric Variable

Occasionally, record level data needs to be examined during baseline analysis. The PROC REPORT procedure can not only create a table of record level data but also compute new variables:

```

TITLE      "Designated Hitter Records";
TITLE2    "PROC REPORT";
PROC REPORT DATA=BB02;
WHERE Group EQ "Designated Hitter";
COLUMNS Name nHits nAtBat ratio pctHts;
  DEFINE Name      / "Name" WIDTH=25;
  DEFINE nHits     / ANALYSIS MEAN "# Hits" WIDTH=25;
  DEFINE nAtBat    / ANALYSIS MEAN "# At Bats" WIDTH=25;
  DEFINE ratio     / COMPUTED "Ratio Hits/At Bats" FORMAT=6.2;
  DEFINE pctHts    / "# Hits" WIDTH=25;
  COMPUTE ratio;
      Ratio = nHits.MEAN / nAtBat.MEAN;
  ENDCOMPUTE;
RUN;

```

Figure 3 displays the output created by the base SAS code. A similar visualization (although lacking the computed variable) can be created in Enterprise Guide by first using the Filter and Sort tab to isolate Designated Hitters records and then using the "List Data" option under the Describe tab to create the report. To fully recreate the output created by PROC REPORT, create the ratio variable prior to using "List Data."

Designated Hitter Records
PROC REPORT

Name	# Hits	# At Bats	Ratio Hits At Bats	% Hits
Thornton, Andre	92	401	0.23	22.9
Baylor, Don	139	585	0.24	23.8
McRae, Hal	70	278	0.25	25.2
Grubb, Johnny	70	210	0.33	33.3
Phelps, Ken	85	344	0.25	24.7
Parrish, Larry	128	464	0.28	27.6
Sheets, Larry	92	338	0.27	27.2
Easler, Mike	148	490	0.30	30.2
Jackson, Reggie	101	419	0.24	24.1
Kittle, Ron	82	376	0.22	21.8
Smalley, Roy	113	459	0.25	24.6

Figure 3. Record Level Reporting and Analysis

COHORT SUCCESS – NUMERIC

Success variables may be numeric or categorical. Figure 3 displayed the ratio of player hits to player at bats. PROC SGPLOT can create a visual representation of the ratio through the use of a scatter plot:

```
TITLE      "Visualization of Hitting Success";
TITLE2    "PROC SGPLOT";
PROC SGPLOT DATA=BB02;
    SCATTER X=nAtBat Y=nHits;
RUN;
```

The PROC SGPANEL procedure has the ability to display multiple plots within a single output. By using PROC SGPANEL, the visualization of hitting success can be broken down by analytical group:

```
TITLE      "Visualization of Hitting Success by Group";
TITLE2    "PROC SGPLOT";
PROC SGPLOT DATA=BB02;
    PANELBY Group;
    SCATTER X=nAtBat Y=nHits;
RUN;
```

Figure 4 displays the output created by the base SAS code. Similar visualizations can be created in Enterprise Guide by using the "Scatter Plot" and "Scatter Plot Matrix" options under the Graph tab.

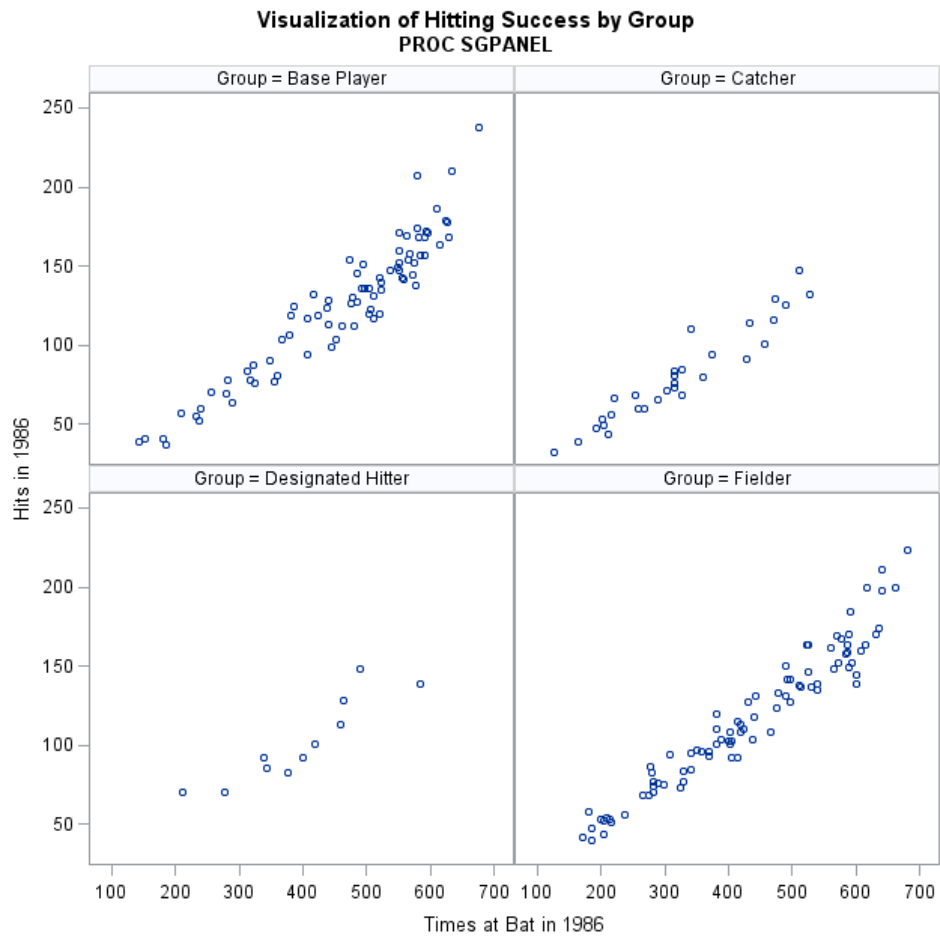
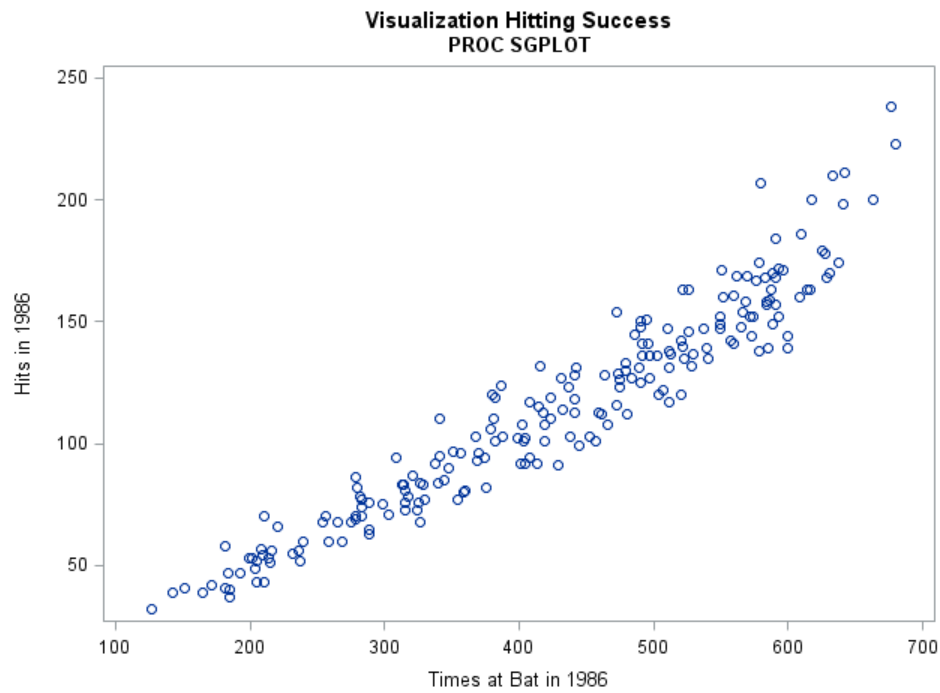


Figure 4. Visualizations of a Numeric Variable by Cohort and Sub-Groups

COHORT SUCCESS –CATEGORICAL

The definition of success can be internally or externally determined. For this analysis, the standard of hitting success is determined by the mean percent of hits by all players in the cohort:

```
TITLE      "Overall Hitting Success";
TITLE2    "PROC MEANS";
PROC MEANS DATA=BB02 NAXDEC=0;
          VAR pctHts;
RUN;
```

Figure 5 displays the output created by the base SAS code. A similar visualization can be created in Enterprise Guide by using the “Summary Statistics Wizard” option under the Describe tab. Once the mean percent is determined the numerical variable pctHts can be coded into the categorical variable Hitting:

```
IF PctHts LT 27 THEN Hitting="Low Success";
ELSE Hitting="Success";
```

Overall Hitting Success PROC MEANS

Analysis Variable : PctHts				
N	Mean	Std Dev	Minimum	Maximum
208	27	3	20	36

Figure 5. Setting a Standard for Comparison

PROC FREQ and PROC SGPLOT can be used to analyze and visualize Hitting:

```
TITLE      "Hitting Success by Group";
TITLE2    "PROC FREQ";
PROC FREQ DATA=BB02;
          TABLES Group*Hitting / NOPERCENT NOCOL;
RUN;
```

```
TITLE      "Hitting Success by Group";
TITLE2    "Success = Hitting 27% or More";
TITLE3    "PROC SGPLOT";
PROC SGPLOT DATA=BB02;
          VBAR Group / GROUP=Hitting;
RUN;
```

Figure 6 displays the output created by the base SAS code. Similar visualizations can be created in Enterprise Guide by using the “Summary Tables Wizard” option under the Describe tab and the “Bar Chart” option under the Graph tab.

Hitting Success by Group
Success = Hitting 27% or More
PROC FREQ

Frequency Row Pct	Table of Group by Hitting		
	Group	Hitting	
		Low Success	Success
Base Player	38 47.50	42 52.50	80
Catcher	26 86.67	4 13.33	30
Designated Hitter	7 63.64	4 36.36	11
Fielder	52 59.77	35 40.23	87
Total	123	85	208

Hitting Success by Group
Success = Hitting 27% or More
PROC SGPLOT

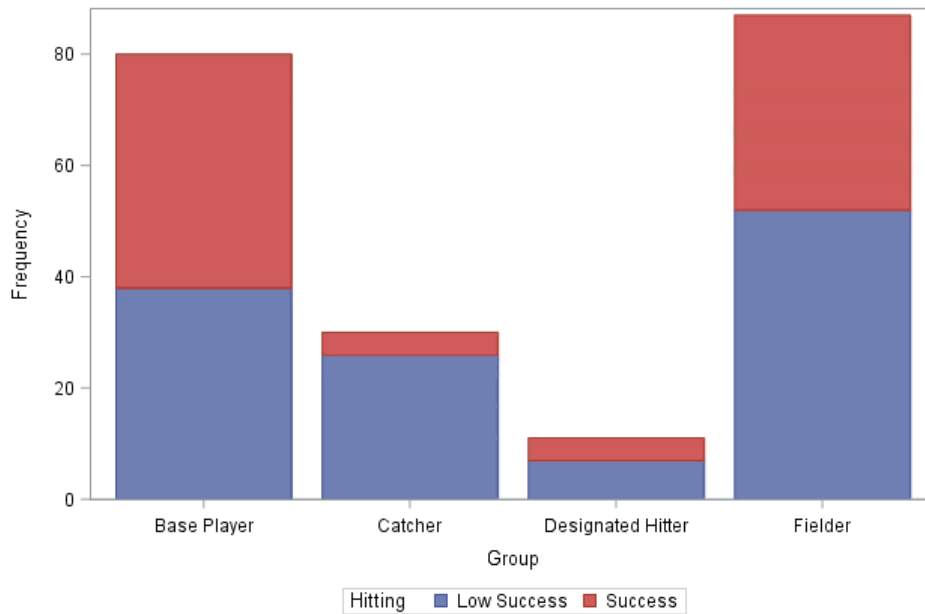


Figure 6. Analysis and Visualization of a Categorical Variable

Frequently, there is a request to display group success rates in comparison to the standard. If Hitting is recoded as 1 and 0 for “Success” and “Low Success” the new variable (HitSucc) can be summarized and an overall percent success by group calculated (SuccessPercent):

```
IF SUBSTR(Hitting,1,1) EQ 'S' THEN HitSucc=1;
ELSE HitSucc=0;
```

Adding a reference line of 27% to a bar chart of SuccessPercent by Group generates the desired chart:


```

TITLE      "Comparison Group Success to Standard";
TITLE2    "PROC SGPLOT";
PROC SGPLOT DATA=BB02;
  VBAR Group / Response=SuccessPercent BARWIDTH=0.5;
  REFLINE 27 / TRANSPARENCY=0
           LABEL="Standard = 27%";
RUN;

```

Figure 7 displays the output created by the base SAS code. A similar visualization can be created in Enterprise Guide by using the “Bar Chart” option under the Graph tab.

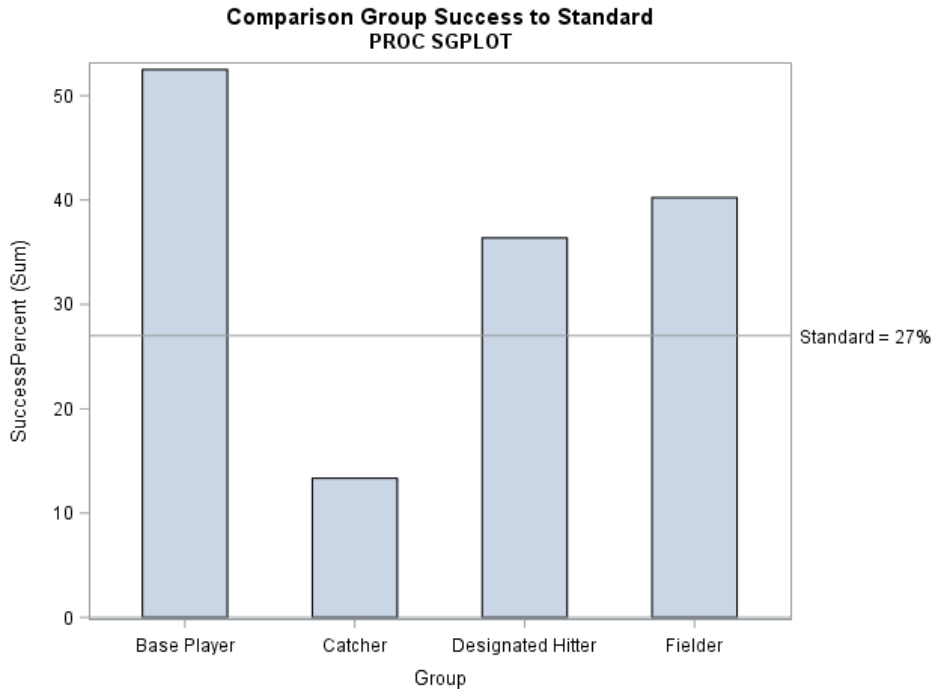


Figure 7. Comparison of Group Success to Standard

CONCLUSION

Institutions of education need a thorough understanding of the current state of student enrollment and success before choosing interventions and policy changes in their work to improve educational equity. SAS provides a wide array of choices for data analysis and visualization; the best choice depends on the question being asked, the type of variable chosen for the analysis, and the audience who will work with the information generated.

REFERENCES

- American Institutes for Research (2020). “Equity in education.” <https://www.air.org/topic/education/equity-education>
- Fink, J. & Jenkins, D. (2020, June). *Unpacking program enrollments and completions with equity in mind*. CCRC Analytics, 1-16. <https://ccrc.tc.columbia.edu/media/k2/attachments/unpacking-programenrollments-completion-equity.pdf>

Organization for Economic Co-Operation and Development (2018). *Equity in education: Breaking down barriers to social mobility*. PISA, OECD Publishing: Paris.
<http://www.oecd.org/pisa/Equity-in-Education-country-note-US.pdf>

Peters, E. (2020, July 28). *IR is a critical partner in diversity, equity, and inclusion work* [interview]. <https://www.airweb.org/resources/publications/eair-newsletter/article/2020/07/28/ir-is-a-critical-partner-in-diversity-equity-and-inclusion-work>

Sargrad, S., Harris, K., Partelow, L., Campbell, N., & Jimenez, L. (2019, July 2). *A quality education for every child: A new agenda for education policy*.
<https://www.americanprogress.org/issues/education-k12/reports/2019/07/02/471511/quality-education-every-child/>

Simmons, D. (2020). *Why Covid-19 is our equity check*. Educational Leadership Special Report. https://www.educationalleadership-digital.com/educationalleadership/april_2020_special_report_free/MobilePagedArticle.action?articleId=1577556#articleId1577556

Slaughter, S. J., & Delwiche, L. D. (2010). *Using PROC SGPLOT for quick, high-quality graphs*. Presented at Western Users of SAS Software 2010, San Diego, California.
https://www.lexjansen.com/wuss/2015/141_Final_Paper_PDF.pdf

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Kelly D. Smith
Senior Research Analyst
Central Piedmont Community College
kelly.smith@cpcc.edu
kds.aewas@gmail.com

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.
Other brand and product names are trademarks of their respective companies.