

In 2018, the Ethnicity/Ancestry Expansion Project began considering a series of changes to the race and ethnicity section in CCCApply that ultimately increased the available response options from 21 to 194 in the hopes of better capturing the “diverse and distinct communities served by the California Community Colleges”<sup>1</sup>. The current response set includes ancestry groups with a population greater than 10,000 in the state of California, plus for Native Americans whose population data are not generally available, the expansion includes all tribes recognized by the state and several of the largest nationally-recognized tribes in the United States. We’ve referred to these additions as expanded ancestry rather than expanded ethnicity, to distinguish it from the racial and ethnic categories that have been the focal point of the college’s ongoing equity work.

With many more ancestry options available to students, we are now seeing student groups we haven’t seen previously and many different ancestry combinations that were not visible when we focused solely on race. For example, we can now see Nicaraguan students for the first time in our data (Figure 1). Eight of the top 20 largest ancestry groups in academic year 2019-2020 are multi-ethnic, with the largest group being Mexican and Other Hispanic or Latino. Overall, in any given academic year, there are nearly a thousand distinct ancestry patterns at PCC. And the number of students within each of these patterns varies widely. In making these patterns visible and known, our office also saw that finding groups within in these patterns would support the PCC community in understanding and using expanded ancestry data.

Because our goal was to supplement, rather than supplant, the data driving PCC’s existing equity work, we modeled our approach toward creating the new categories after the procedure used to create the federally-recognized race category, known locally as IPEDS race<sup>2</sup>. Our models found nine major groups or clusters in the data (Figure 2). We

---

<sup>1</sup> <https://cccnex.t.jira.com/wiki/spaces/PD/pages/821887199/Ethnicity+Ancestry+Expansion+Project>

<sup>2</sup> First, we divided 21 race ethnicity response choices into the seven pre-defined groups, which IPEDS then rolls up into a singular race designation per student. Mimicking this, a principal components analysis was used to group the 194 expanded ethnicity/ancestry responses. An artificial neural network analysis was then used to create an expanded ethnicity/ancestry-based analog of the seven pre-defined groups traditionally used to create IPEDS race. These two steps gave us a way of consolidating the

included IPEDS Race categories into our analysis with the understanding that while they do obscure nuanced ancestries, they are used by other colleges where our clusters may not apply. Mapping back to IPEDS also lets us roll the data back up if we need to for standardized reports and benchmark our progress over time. Within three clusters, we noted several branches that map back to different IPEDS Race groups and we recommend be regarded as unique, emerging subsets of the IPEDS Race groups each of them falls under. To emphasize and reinforce this distinction, we labeled those clusters with the first two letters of the IPEDS Race group to which they map (Figure 3).

---

194 responses into more manageable chunks. Proceeding to classifying students, a hierarchical cluster analysis (Kohonen) was used to fuse the predicted probabilities from the principal components analysis and the artificial neural networks analysis, suggesting that we might be able to identify 12 types of students based on their combination of expanded ethnicity/ancestry responses. A second, final hierarchical cluster analysis (Kohonen) was then used to combine the first cluster analysis's results with IPEDS race in order to interface these common 12 socio-politically agnostic ancestry patterns with the college's ongoing equity work and empowerment programs centered around IPEDS race ids.

Expanded Ancestry	≡	Number of Students	Rank by Number of Students	Retention Rate	Success Rate
Mexican;Other Hispanic or Latino		6,775	1	82.0%	66.6%
Other White		4,731	2	87.2%	76.4%
Chinese		4,573	3	89.6%	77.6%
Unknown or Not Reported		3,275	4	88.1%	72.9%
Other Hispanic or Latino		1,849	5	82.6%	60.7%
Mexican;Other Hispanic or Latino;Other White		1,789	6	83.3%	70.5%
Other Asian		1,671	7	88.9%	77.8%
African American		1,174	8	83.3%	64.0%
Nicaraguan;Other Hispanic or Latino		1,076	9	81.0%	65.0%
Mexican		1,060	10	77.2%	57.1%
Filipino		1,026	11	85.5%	75.5%
Vietnamese		693	12	89.0%	80.4%
Korean		615	13	89.6%	81.1%
Chinese;Vietnamese		475	14	89.9%	81.6%
Mexican;Nicaraguan;Other Hispanic or Latino		451	15	81.2%	66.1%
Other Hispanic or Latino;Other White		449	16	82.9%	68.2%
Chinese;Other Asian		397	17	87.9%	80.9%
Nicaraguan		377	18	81.1%	59.3%
Indian		189	19	86.9%	76.4%
Mexican;Other White		179	20	81.1%	65.5%
Others		7,133	21	83.6%	70.7%
<b>Grand Total</b>		<b>39,957</b>		<b>85.0%</b>	<b>71.4%</b>

Figure 1 The top 20 expanded ancestries and their retention and success rates for academic year 2019-2020.

With cluster analysis, we used IPEDS, Expanded Ancestry, and group size to create clusters.

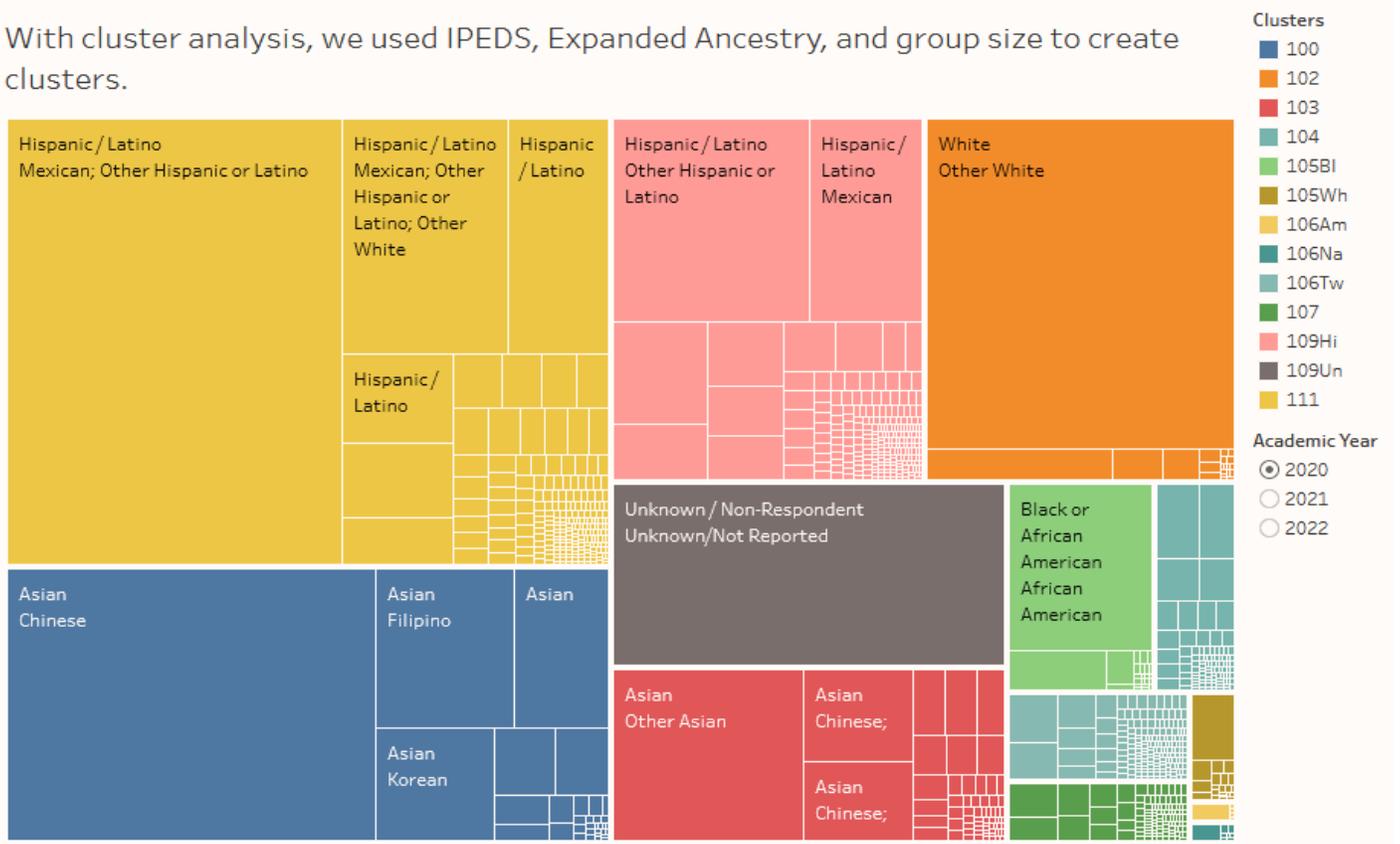


Figure 2 Nine major clusters, with several branches denoted with the first two letters of their corresponding IPEDS Race category. The size of each rectangle indicates the size of the group.

Based on our analysis, we’re provisionally characterizing the clusters and branches as either established or emerging. Established indicates groups whose race and ethnicity data have been traditionally available to our campus, for example, Mexican and Hispanic or Latino ancestries in cluster 111. Emerging denotes groups whose ancestries we are now seeing with the availability of nuanced ancestry data, for example, Armenian and Other White in cluster 105Wh. Over time as more students report expanded ancestry data, clusters that are currently established may change, for example, clusters 106Am and 106Na. Due to these changes, we have chosen to keep the clusters unnamed and provide the top five ancestry patterns in each cluster as a working description (Figure 4). Some groups e.g. Other Asian in cluster

107, may be the result of students selecting expanded ancestry options that do not map back to their historic IPEDS category.

IPEDS Race Category	Cluster Number	Branch	Label	Established or Emerging?
American Indian	106	Am	106Am	Established
Asian	100	--	100	Established
Asian	103	--	103	Emerging
Black or African American	105	Bl	105Bl	Established
Hispanic or Latino	107	--	107	Emerging
Hispanic or Latino	109	Hi	109Hi	Emerging
Hispanic or Latino	111	--	111	Established
Native Hawaiian and Other Pacific Islander	106	Na	106Na	Established
Two or More	104	--	104	Established
Two or More	106	Tw	106Tw	Established
Unknown or Not Reported	109	Un	109Un	Established
White	102	--	102	Established
White	105	Wh	105Wh	Emerging

Figure 3 Describing cluster labels.

Who's in these clusters? Below we list the top five groups.

IPEDS Race	Clusters	Expanded Description List Display	Number of Students	Rank by Number of Students
American Indian or Alaska Native	106Am	Other American Indian	39	1
		Other Hispanic or Latino; Othe..		5
		Navajo; Other American Indian		4
		Chippewa; Elem Indian Colony ..		3
		Aztec or Mixtec; Other Americ..		2
Asian	100	Chinese	4,567	1
		Filipino	1,017	2
		Vietnamese	689	3
		Korean	613	4
		Indian	186	5
	103	Other Asian	1,506	1
		Chinese; Vietnamese	475	2
		Chinese; Other Asian	396	3
		Cambodian	99	4
		Other Asian; Vietnamese	95	5
Black or African American	105BI	African American	1,128	1
		African American; Other Nort..	179	2
		Other North African	42	3
		Subsaharan African: Ethiopian..		4
		Subsaharan African: Other Su..		5
Hispanic / Latino	107	Peruvian	79	1
		Other White	56	2
		Mexican; Other Hispanic or La..	51	3
		Other Asian	39	4
		Other Hispanic or Latino; Filipi..	31	5
	109Hi	Other Hispanic or Latino	1,829	1
		Mexican	1,060	2
		Other Hispanic or Latino; Othe..	433	3
		Peruvian; Other Hispanic or La..	241	4
		Mexican; Other Hispanic or La..	226	5
	111	Mexican; Other Hispanic or La..	6,775	1
		Mexican; Other Hispanic or La..	1,789	2
		Nicaraguan; Other Hispanic or..	1,076	3
		Mexican; Nicaraguan; Other Hi..	451	4

Figure 4 A snapshot view showing the Top 5 ancestry patterns ('Expanded Description List Display') in each cluster and branch for academic year 2019-2020. The ranking is based on the number of students and blanks denote a group that has fewer than 10 students.

## Retention, Success, and Completion Outcomes

IPEDS Race	Clusters	Number of Students	Enrollments	Retention Rate	Success Rate	AA AS Completed	ADT Completed	Completed Cert
American Indian or Alaska Native	106Am	43	186	75.8%	58.6%	3	2	0
Asian	100	7,474	39,848	88.9%	77.9%	1,247	476	113
	103	3,108	13,593	89.1%	79.1%	346	130	39
Black or African American	105Bl	1,391	6,352	82.7%	63.5%	99	49	21
Hispanic / Latino	111	12,202	62,884	82.0%	67.1%	1,417	752	191
	109Hi	5,125	22,464	81.4%	63.4%	409	204	75
	107	493	1,937	84.3%	68.9%	58	15	7
Native Hawaiian or Other Pacific Islander	106Na	41	207	87.0%	72.9%	0	0	0
Two or More Races	104	765	3,820	84.6%	72.6%	94	43	12
	106Tw	724	3,151	82.9%	71.0%	60	24	5
Unknown / Non-Respondent	109Un	3,282	11,955	88.1%	72.9%	282	108	24
White	102	5,077	23,479	87.2%	76.5%	540	249	72
	105Wh	232	838	86.2%	74.7%	6	1	1
<b>Grand Total</b>		<b>39,957</b>	<b>190,714</b>	<b>85.0%</b>	<b>71.4%</b>	<b>4,561</b>	<b>2,053</b>	<b>560</b>