



Evidence-to-Impact
Collaborative

Analysis of Capital Access for Entrepreneurs of Color

By Yoon Sun Hur



PennState
College of Health and
Human Development



THE
**VENTURE
EQUITY**
PROJECT

OVERVIEW

What makes an effective data system?

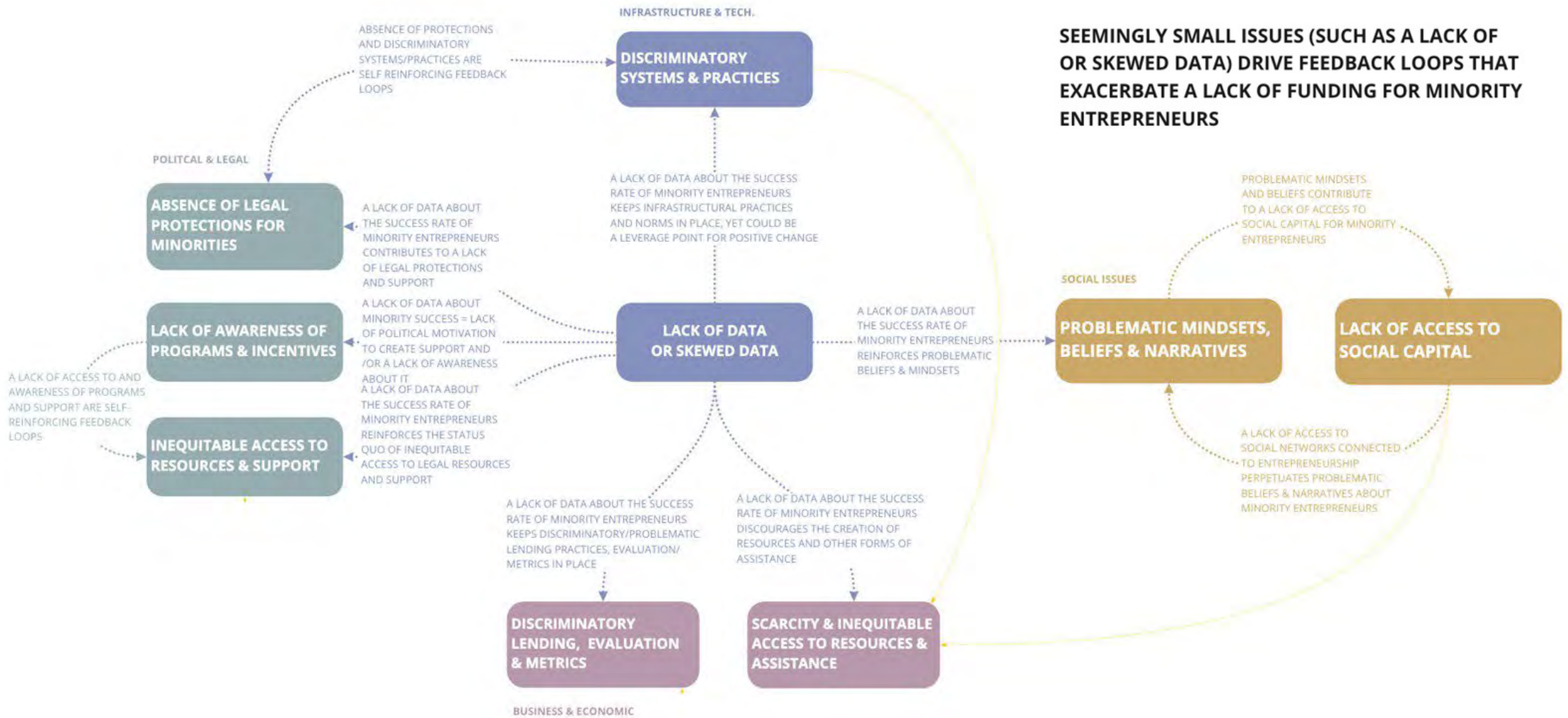
What data streams are needed?

What data do we have?

What does the data tell us?

What is next?





STUDYING THE CAPITAL ACCESS ECOSYSTEM

Insights from the Transition Design workshops

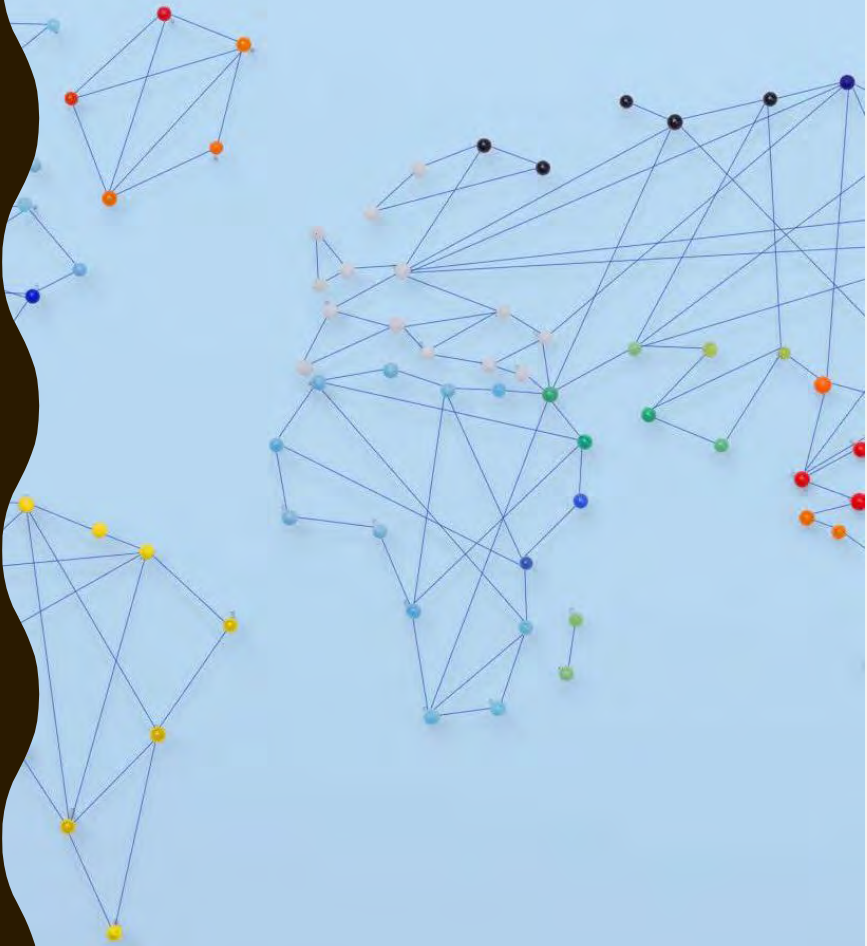
The Need for an Integrated Data System

- A comprehensive view of the larger system that centers entrepreneurs and their businesses
- Data is an essential component of the larger ecosystem for:
 - Triangulating inequities
 - Mapping systemic and structural problems
 - Tailoring interventions
 - Benchmarking success



The Data System

- An integrated data environment comprised of key data and indicators from across the entrepreneurial capital ecosystem
- **UNDERSTANDING THE PROBLEM**
 - Track the magnitude and nature of the inequity in amount of funding going to minority- vs. non-minority-led start-ups
 - Understand inequity within different funding sources and industries
- **BENCHMARKS FOR SUCCESS**
 - Map initiatives and efforts to reduce inequity onto historical, current, and future data to understand whether inequity is being reduced



Only as Good as the Data that Feeds It



A successful data system is fed by varied and complementary data streams



Poor or incomplete data from any stream will jeopardize the index as a whole and undercut confidence in its use



We begin by conceptualizing what this system looks like in an open data environment



Then we consider a data system in the current environment

Core Components of an Effective Data System

Equity-oriented Map of the Funding Ecosystem

- Capital stacks of entrepreneurs - historical and current
- Diversity of investor portfolios
- Pre-seed funding sources and levels
- Segmented by sector and race

Metrics & Indicators

- Funders' success or challenges supporting entrepreneurs of color
- Entrepreneurs of color experiencing equity / inequity with their white peers

What Data is Needed?

Racial and ethnic
identity of
entrepreneurs

Company
descriptives

Pre-seed and seed
funding

Funding
rounds

Government
programs &
support
(local, state, and
federal levels)

Accelerator and
incubator inputs

Data on Hand

(Quantitative & Qualitative)

Subgrantee	Description of Data	Contact
Camelback	<ul style="list-style-type: none">- Application for 2021 (observations [obs.] #304) and 2022 R1 (obs. #186) uploaded- Race, gender of applicants, funding stage of applicant's company information included	Aaron Walker (aaron@camelbackventures.org)
Extended Venture	<ul style="list-style-type: none">- Individual and company information (race/gender/funding)- 8-month time interval	Erika Brodnock (Erika@extend.vc (Erika@kami.coach)
ParentPreneur	<ul style="list-style-type: none">- Video files (30) // 135 grant applicant directory- Directory includes basic demographics (first/last name, company name)- No information about race/gender and funding stage of company	James Oliver (james@parentpreneurfoundation.org)
VetsinTech	<ul style="list-style-type: none">- 6 VEP-related event registration files- First name, last name, email, age, etc.- No company name	Ikram Mansori (imansori@vetsintech.co)
Institute for Entrepreneurial Leadership (IEFL)	<ul style="list-style-type: none">- Success Circles program information // WOCCON event information	Jill Johnson (jill@weareifel.org)
Capital Enterprise	<ul style="list-style-type: none">- OneTech file- company name, race, age, gender, disability, LGBT information included	Ana Lucia Buckman (analucia@capitalenterprise.org)

Other Data

NEC event registration data	<ul style="list-style-type: none">- Participants of NEC programs and events in 2021-2022- Demographics (race, gender, age, role, etc.) of entrepreneurs.	
-----------------------------	---	--

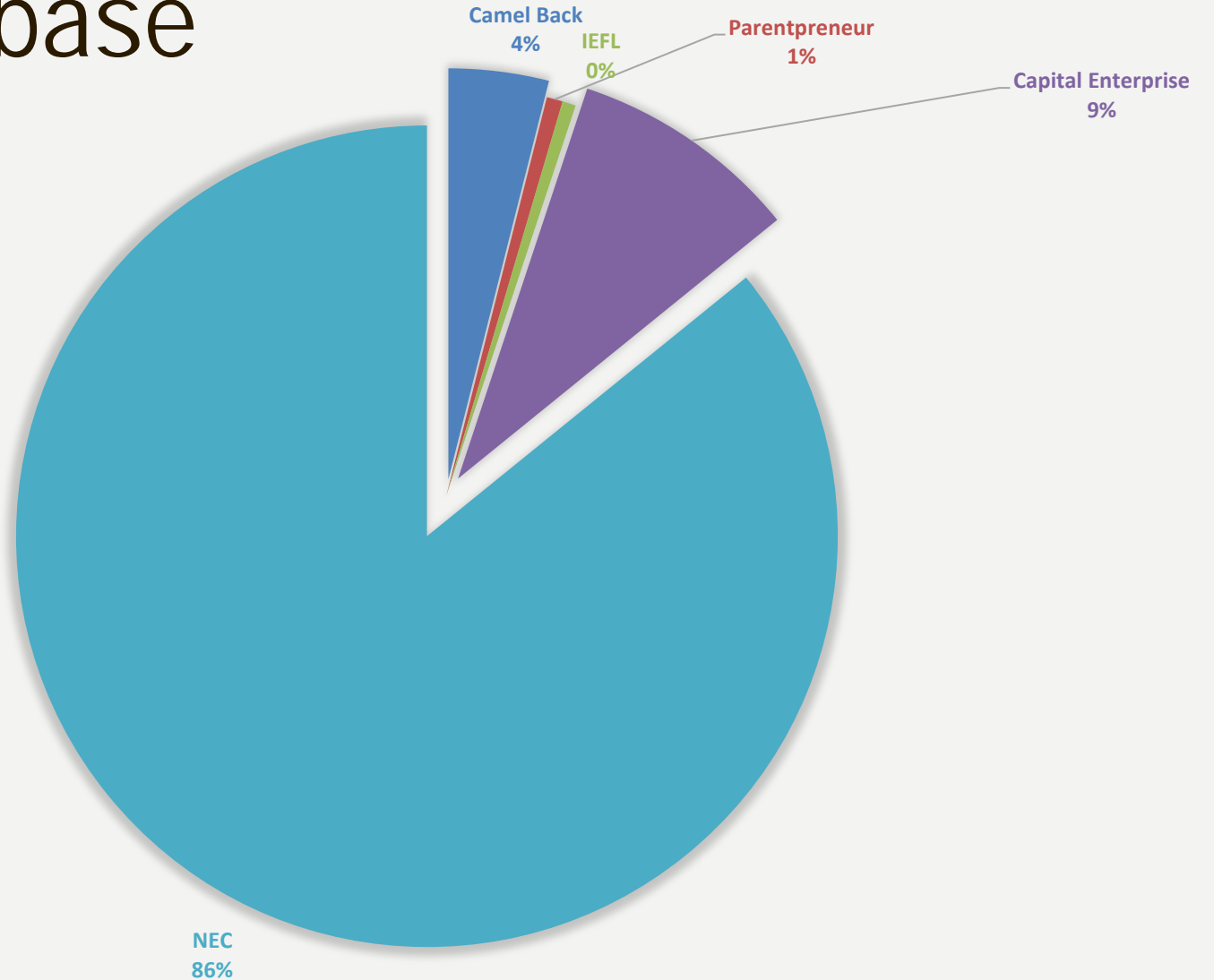
Building the Data System Using Crunchbase

Why Crunchbase?

- Since 2005, Crunchbase has been a leading provider of data on entrepreneurs' activity. Its data is widely used by industry and academics.
- The subgrantee data provides abundant demographic information about entrepreneurs (gender and race), and some basic information about their funding. But there is little to no information about funding details (e.g., funding type, funders, funding stage, funding history, etc.)
- We linked the subgrantee data to Crunchbase, a database operated by the media company Tech Crunch that contains detailed information on startups including their funding information as collected from new publications, venture capital firms, executives, entrepreneurs, and investors.
- Under a partnership between the NEC, PSU, and Crunchbase, Crunchbase provided full access to Crunchbase API.

Linking to Crunchbase

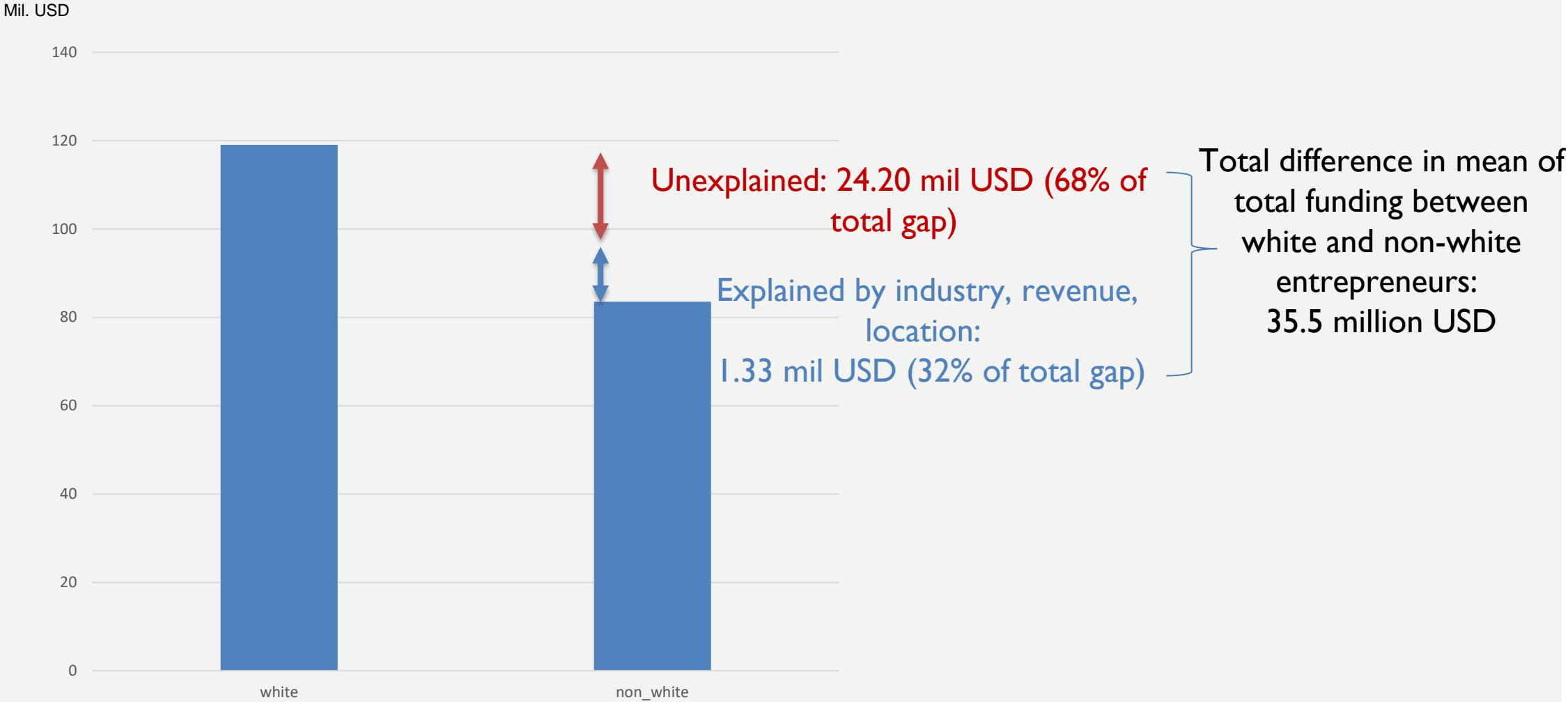
- Of the 5 subgrantees whose data connected to Crunchbase, 3,023 matched
- Of these 3,023 companies -
 1. 86% from NEC data
 2. 9% from Capital Enterprise data
 3. 4% from Camelback data
 4. 1% from ParentPreneur data
 5. Less than 1% from IFEL data



Explaining Funding Gaps: Methods

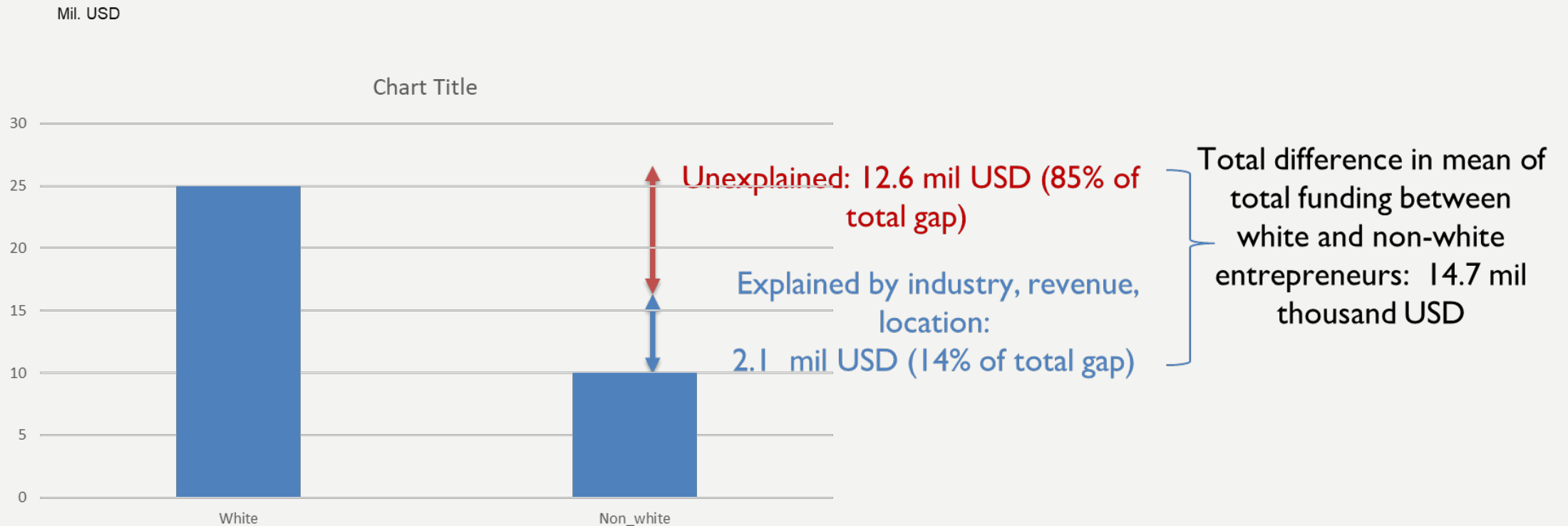
- We introduced the Blinder-Oaxaca decomposition method to be applied in explaining inequality in the entrepreneurial outcome across race/ethnicity.
- The Blinder-Oaxaca decomposition method is one of the most widely used methods for decomposing mean differences in specific outcomes between 2 groups. It is first applied to explain the gap in mean wages between women and men, and has been applied toward explaining various outcomes in the field of labor economics and education/health economics.
- The difference between the observed white entrepreneurs' funding distribution ($F_{Y(W|W)}$) and the observed non-white entrepreneurs' funding distribution ($F_{Y(NW|NW)}$) can be decomposed as follows:

Explaining Funding Gaps



(Sample: NEC community & Crunchbase, n=2,592)

Explaining Funding Gaps- early stage only



How do we use different data?

In summary:

- We were able to use NEC & Camelback data in a statistical analysis
- We were able to use ParentPreneur, Extend VC, VetsinTech, and Capital Enterprise data sets in a descriptive analysis

Why is this?

- The sample size & consistency across some data sets prevented use in a statistical analysis
- Funding and revenue information (vital to connecting to Crunchbase) was not provided by most subgrantees
- **However**, these descriptions of each subgrantee's data sets can help us hypothesize about the unexplained gap between funding for white and non-white founders/companies. For example:
 - How the market and product is seen and valued by investors
 - The relationship between the location of a company and locations of investors

Strengthening the Data System

Year 1 feasibility work resulted in the following actionable steps for Year 2:



1. Increase the capacity to harmonize partner data by adding standardized items across systems based on common constructs derived from qualitative analyses



2. Leverage new insights about data gaps based on the lived experiences and inferential analysis to ingest and connect additional external data sets (e.g., Census, Gust, SEC)



3. Continue to support partners in additional data collection efforts

Next Steps

A number of opportunities have emerged to greatly enrich the data being collected and analyzed for this project. Potential collaborators identified by the Nasdaq Entrepreneurial Center have offered initial insights into data sets under development and/or in late stages and ready for analyses that will add to the robust nature of the data set. Increasing the number of data sets introduced into this Index will provide greater insights into whether inequity is being reduced, and if so, how/in what ways this is occurring so that the lessons may be applied more broadly and evenly.

Recommendations for Year 2 Data Collection and Collaboration Opportunities are as follows:

1. Hold in-depth meetings with the following organizations to explore their data:
 - a. Gust
 - b. StartOut
 - c. Diversity VC
 - d. Small Business Majority
 - e. Angel Research Institute
 - f. University of New Hampshire
 - g. Babson College / Diana Project

2. Request and explore additional federal data sets, with access requiring form submissions to federal agencies
 - a. Securities and Exchange Commission (SEC) Form D
 - b. U.S. Census data (additional elements)

Analysis of Capital Access for Entrepreneurs of Color

Appendix



Evidence-to-Impact
Collaborative

THE
**VENTURE
EQUITY**
PROJECT



- The difference between the observed White entrepreneurs' funding distribution ($F_{Y(W|W)}$) and the observed Non-white entrepreneurs' funding distribution ($F_{Y(NW|NW)}$) can be decomposed as follows:

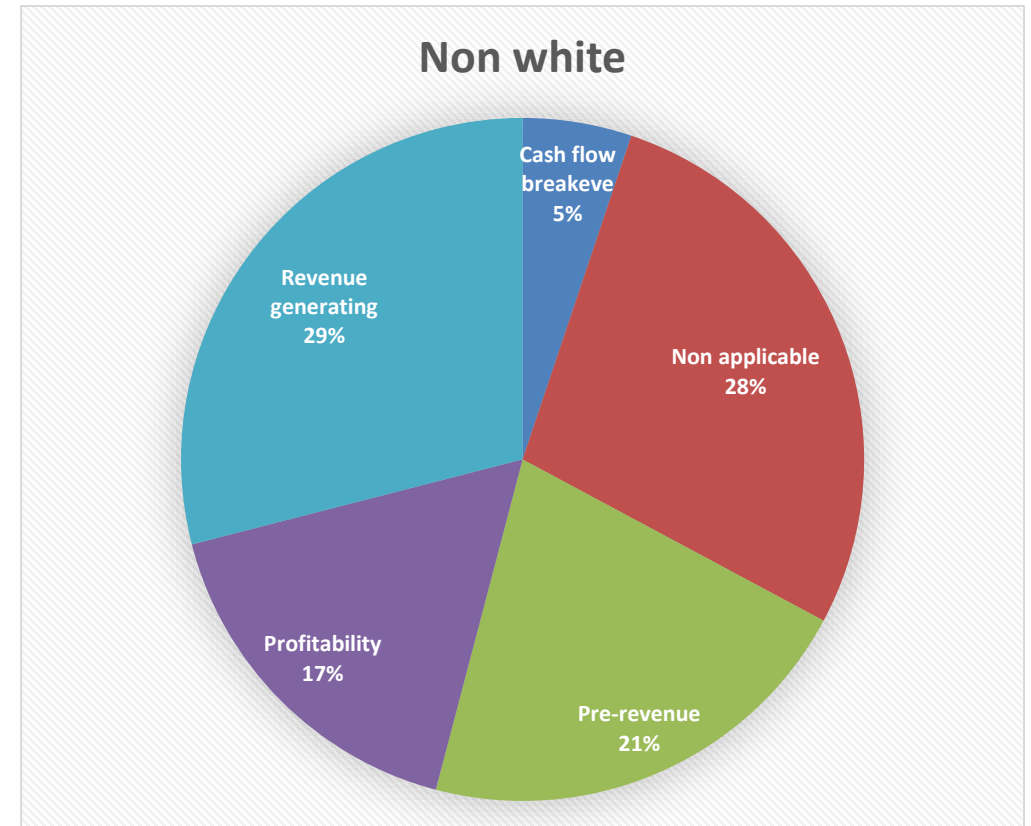
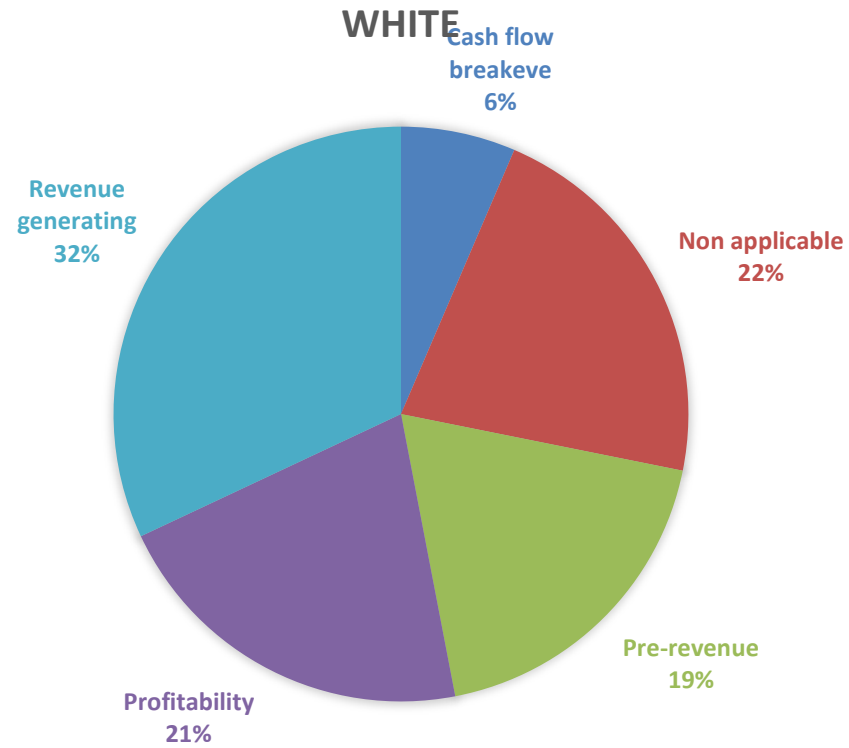
$$F_{Y(W|W)} - F_{Y(NW|NW)} = \underbrace{[F_{Y(W|W)} - F_{Y(W|NW)}]}_{\text{Characteristic effect}} + \underbrace{[F_{Y(W|NW)} - F_{Y(NW|NW)}]}_{\text{Unexplained Effect}}$$

Characteristic effect
(Industry, revenue level,
location)

Unexplained Effect

Blinder – Oaxaca Equation

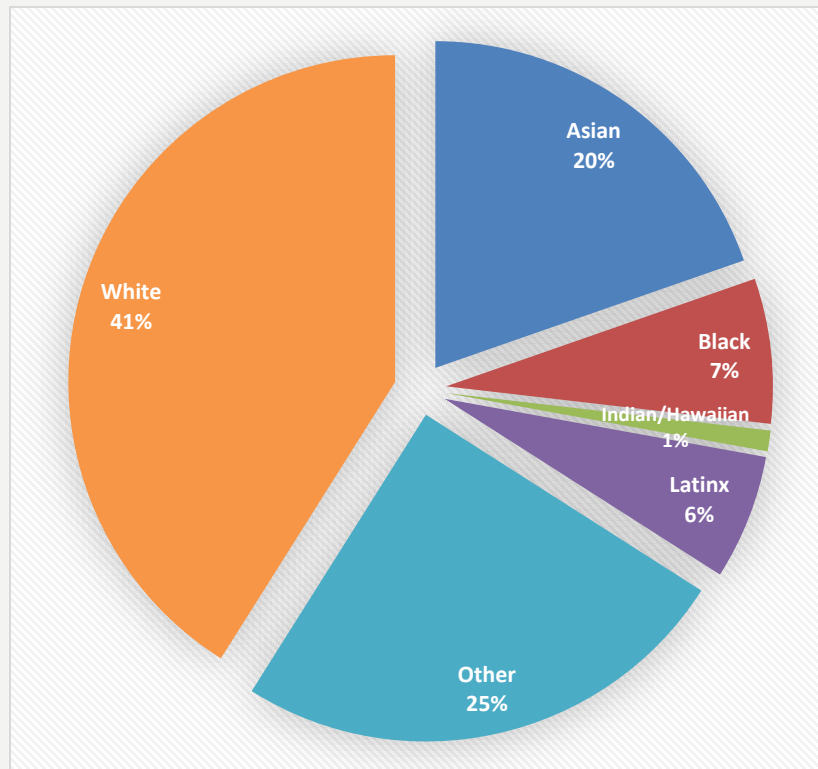
Revenue



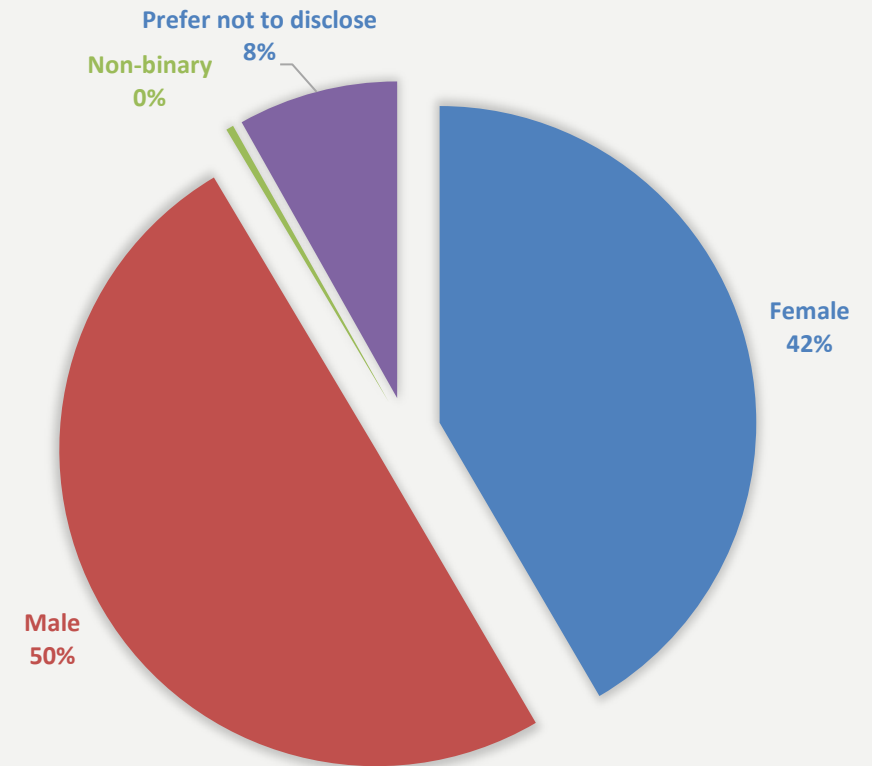
Appendix B: Going Deeper

Demographic Information on Entrepreneurs

Race



Gender



(Sample: NEC community & Crunchbase, n=2,592)

Demographic Information on Entrepreneurs

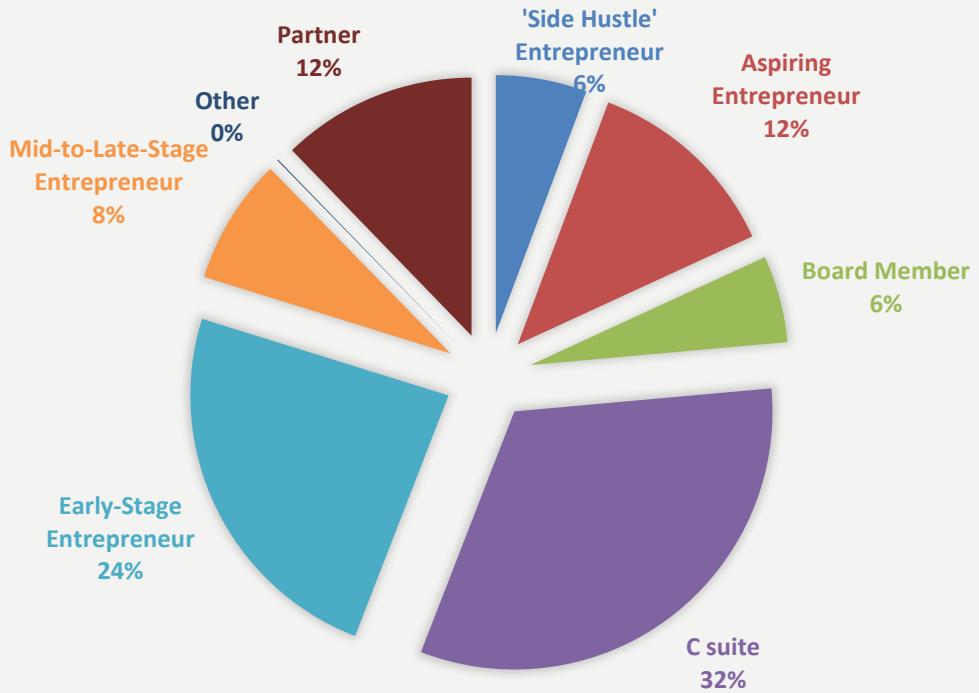
Breakdown by Race & Ethnicity

	Female	Male	Non-binary	Prefer not to disclose	Total
Asian	223	254	0	3	480
Black	111	64	0	0	175
Indian/Hawaiian/ Pacific Islander	9	15	0	1	25
Latinx	71	77	1	2	151
Other	169	229	5	205	608
White	425	575	1	1	1001
Total	1008	1214	7	212	2440

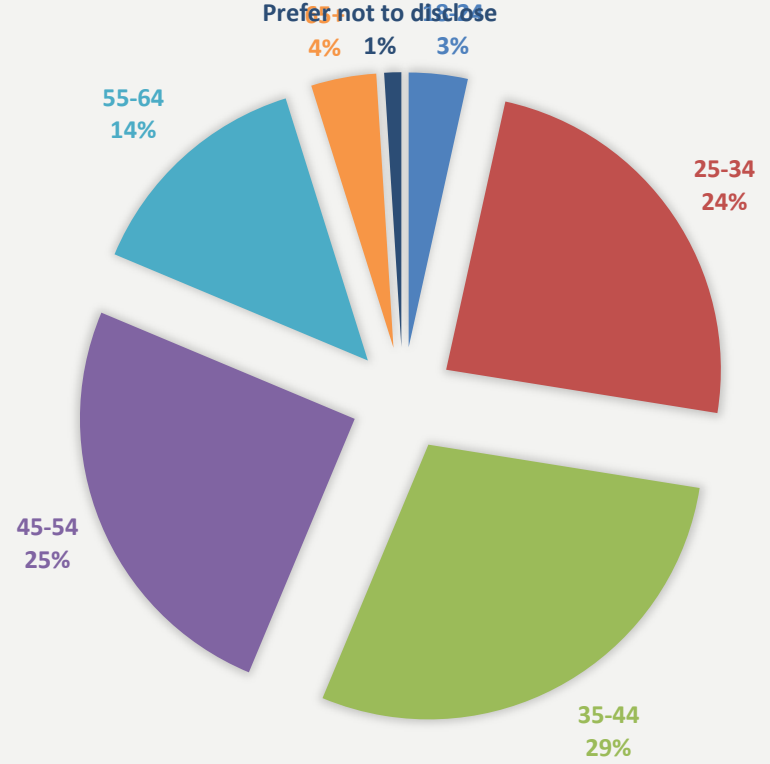
(Sample: NEC community & Crunchbase)

Demographic Information on Entrepreneurs

Role in the Company



Age



(Sample: NEC community & Crunchbase, n=2,592)

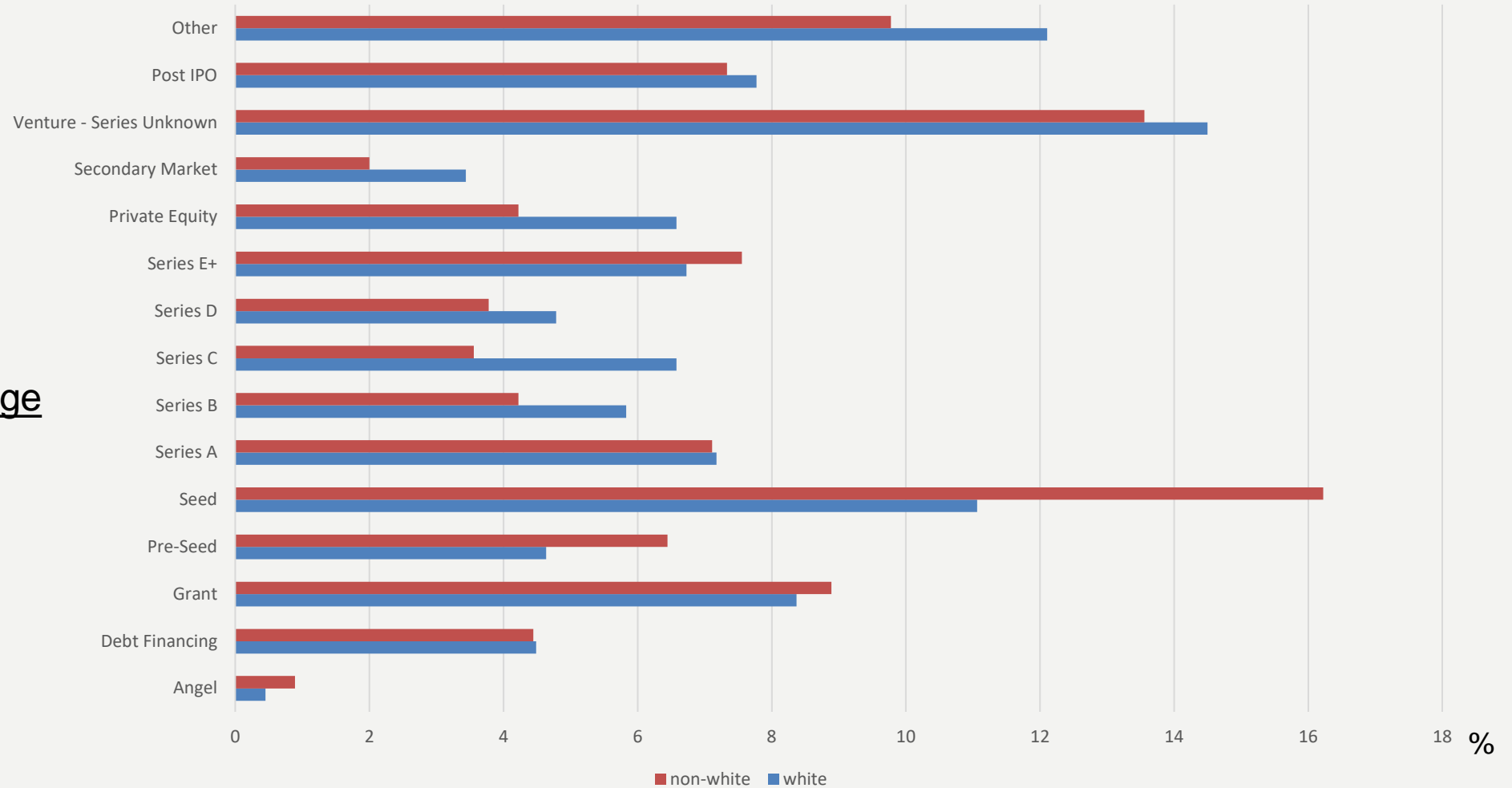
Company Sectors

	White	Non white
Technology	466	358
Finance	216	156
Other	173	187
Healthcare	156	85
Education	58	42
Legal, Insurance	54	35
Business, Information	51	33
Marketing, PR, Media	44	26
Health, Beauty, Fitness	39	31
Hospitality, Food & Beverage	35	35
NGO	31	5
Real Estate	29	25
Retail	29	24
Construction, Utilities	25	27
Travel & Transportation	21	17
Personal Services: Home Goods, Services	13	4
Leisure: Gaming, Sports, Events	12	12
Natural Resources, Environmental	12	3
Agriculture ,Forestry, Wildlife	11	8
Government, Public Services	10	13
Safety & Security	3	6

(Sample: NEC community & Crunchbase, n=2,592)

Company Funding Stages

Funding Stage



* Non-white includes Black, Indian/Hawaiian/Pacific Islander, Latinx, and other.

(Sample: NEC community & Crunchbase, n=2,592)

Company Funding Type

(by race)

Last Funding Type	Asian	Black	Indian/Hawaiian /Pacific Islander	Latinx	Other	White	Total
Angel	2	1	0	2	0	1	6
Convertible Note	3	0	0	1	5	8	17
Corporate Round	2	0	1	0	5	6	14
Debt Financing	4	2	0	3	13	26	48
Equity Crowdfunding	1	0	0	1	0	2	4
Grant	18	9	1	2	18	38	86
Initial Coin Offering	3	0	0	0	0	0	3
Non-equity Assistance	5	6	0	3	5	4	23
Post-IPO Debt	4	1	0	1	2	6	14
Post-IPO Equity	14	2	0	4	22	28	70
Post-IPO Secondary	1	0	0	0	0	2	3
Pre-Seed	15	4	1	9	11	16	56
Private Equity	7	1	1	3	13	37	62
Secondary Market	10	2	0	1	5	14	32
Seed	28	7	4	15	32	46	132
Series A	24	5	2	1	21	24	77
Series B	17	1	0	2	14	25	59
Series C	17	2	1	1	12	27	60
Series D	8	3	1	2	7	25	46
Series E	13	0	0	3	7	16	39
Series F	5	0	0	1	9	9	24
Series G	1	1	0	0	6	2	10
Series H	0	0	0	0	1	0	1
Undisclosed	4	1	0	2	5	8	20
Venture - Series Unknown	30	2	1	11	39	65	148
Total	236	50	13	68	252	435	1054

Going Deeper

1) Defining the gaps in outcomes (funding amount, funding type, revenue) across race/ethnicity

2) Examining how much of the gap in outcomes can be explained by differences in observable characteristics including Industry, location of company, Individual characteristics (education, experience) of entrepreneurs

3) Estimating the remaining inequality in outcomes (unexplained gaps in outcomes). This can be due to differential effects of characteristics, maybe discrimination.

Key Points from Data

1. Connections between subgrantee qualitative data & transition design findings

The Parentpreneur Foundation – out of 116 open-ended responses, we identified four themes of challenges raising capital:

- Barriers to resources & access
- Financial barriers
- Barriers in context of capital enterprise and market
- Barriers in social context

Institute for Entrepreneurial Leadership:

- Over half of respondents cited that the value they bring to their audience is the perspective and amplification of groups of people that may not have access or representation otherwise
- Top three challenges prior to COVID-19 were related to lack of resources or knowledge to further business and gain funding: (1) Marketing business, (2) growing sales & increasing revenue, (3) a knowledge gap & knowing what to do to grow business

Transition Design Institute at Carnegie Mellon University:

- “Key groups (investors, stakeholders) don’t recognize the demand for funding for minority entrepreneurs” –Anonymous community organizer
- “There is a lack of transparency when it comes to lending and how banks process loan applications” –Anonymous Black entrepreneur
- “The racism is in risk, and the majority of financial products use credit history and existing assets to make assessments based on financial considerations that are the result of social, political and other forms of discrimination” –Anonymous community organizer

Key Points from Data

2. Industries & sectors represented among subgrantees are chronically undervalued and not funded

- Parentpreneur: Top 3 sectors
 - Health, beauty, and fitness
 - Retail
 - Education
- IFEL: Top 3 sectors & revenue status
 - Retail – 21 out of 32 companies made between \$0 - \$50,000
 - Professional services – 10 out of 23 companies made between \$0 - \$50,000
 - Food service/catering - 11 out of 16 companies made between \$0 - \$50,000
- Transition Design Institute:
 - “A lot of businesses started by minority entrepreneurs tend to be in industries that are not fast growing and not considered attractive by traditional VC – for [example] food, service industries” –Anonymous US capital allocator
 - “Black and brown communities are viewed as the cheap labor of the economy, not as communities worthy of economic development and generation” –Anonymous US community organizer
 - “[There is a] lack of understanding of the underserved communities being served by the entrepreneurs’ businesses” –Anonymous UK capital allocator
- Why does this matter?
 - The founders in these communities are still struggling for funding in their undervalued businesses, even with access to a large organization

Key Points from Data

3. Location has an impact on demographics and opportunities

- Extend VC captured founding team demographics of almost 400 companies across the US. Founding teams that were all female and mixed races were only represented in the Pacific and South Atlantic regions
 - However, at least 30% of every team from every region were all male and all white
- A large majority of VetsinTech network is located on the East Coast
- Transition Design findings:
 - “In some places, the cost of living prohibits people from taking the necessary risks to be successful entrepreneurs” –Anonymous US Black entrepreneur
 - “Gentrification pushed out low income communities of color from geographic locations that became/are becoming tech hubs that easily attract capital (e.g. Silicon Valley/Oakland, New York City/Brooklyn, etc.)” –Anonymous US capital allocator
 - “Minorities disproportionately reside in areas with lack of access to adequate healthcare, food, housing & in some cases broadband access, making it difficult to obtain basic resources necessary to obtain the basics required for a decent quality of life” –Anonymous US community organizer
- Why does this matter?
 - Networks for venture capital are often located in major cities (New York, San Francisco)
 - By being outside of these hubs, entrepreneurs of color may have limited access to these networks that are vitally important

Key Points from Data

4. State and type of data provided

- Least information regarding funding and revenue
 - This information is key to finding company's information in Crunchbase
- Multiple subgrantees' data sets could not be included in an inferential analysis, because:
 - The variables collected were very different across each organization
 - Some of the variables were not in a form that could be analyzed traditionally (for example, open-ended and audio-transcribed data)
- The representativeness of race & ethnicity varied across subgrantee groups
 - IFEL, ParentPreneur, and Capital Enterprise were over-representative of Black and African-American founders & companies
 - Extend VC and VetsinTech were over-representative of white founders and companies
 - No subgrantees had significant representation of Latinx or Pacific Islander founders and companies
- Certain characteristics that provide more information are not consistently asked
 - VetsinTech & IFEL were the only subgrantees to ask about Veteran status
 - Capital Enterprise was the only subgrantee to ask about disability & LGBTQ+
 - Capital Enterprise data reported that white founders had the most identification in the LGBTQ+ community, even with one of the least prevalent ethnicities
- Why does this matter?
 - By identifying what variables exist among smaller data sources, we can work to standardize questions among organizations to create a stronger data environment
 - We must understand how to accurately capture self-reported variables that may be stigmatized in some communities

Linking to Crunchbase

Subgrantee	Description of Data	Link to Crunchbase	Math Rate
Camelback	<ul style="list-style-type: none"> - Application for 2021 (observations [obs] #304) and 2022 R1 (obs. #186) uploaded - Race, gender of applicants, funding stage of applicant's company information included 	119 out of 490	24%
Extended VC	<ul style="list-style-type: none"> - Individual and company information (race/gender/funding) available - 8-month time interval 	Could not link to Crunchbase (no company name)	0%
ParentPreneur	<ul style="list-style-type: none"> - Video files (30) // 135 grant applicant directory - Directory includes basic demographics (first/last name, company name) - No information about race/gender and funding stage of company 	19 out of 135	14%
VetsinTech	<ul style="list-style-type: none"> - 6 VEP-related event registration files - First name, last name, email, age etc. - No company name 	Could not link to Crunchbase (no company name)	0%
IEFL	<ul style="list-style-type: none"> - Success Circles program information // WOCCON event information 	16 out of 143	10.3%
Capital Enterprise	<ul style="list-style-type: none"> - OneTech file - company name, race, age, gender, disability, LGBT information included 	274 out of 575	47%
NEC event registration data	<ul style="list-style-type: none"> - Participants of NEC programs and events in 2021-2022 - Demographics (race, gender, age, role, etc.) of entrepreneurs. 	2,595 out of 8,213	31%

The match rate to Crunchbase of entrepreneurs in the subgrantee and NEC communities varies by subgrantee

Linking to Crunchbase

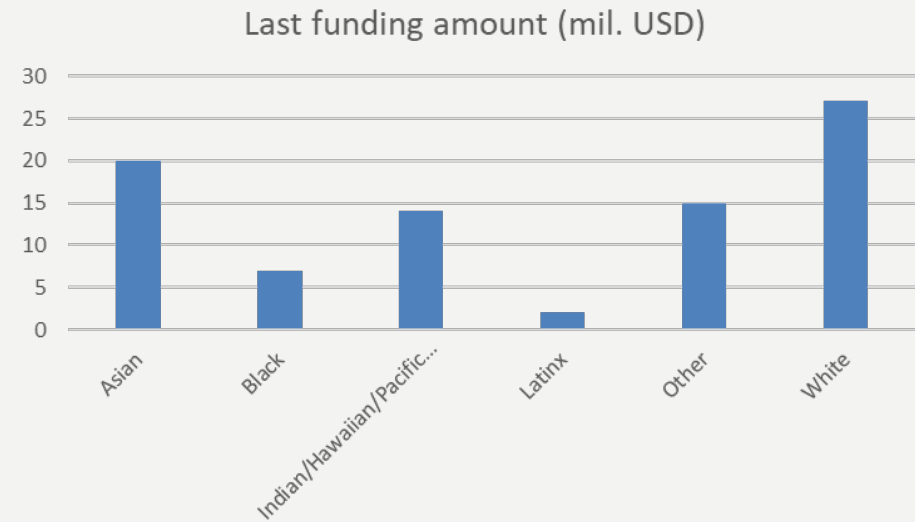
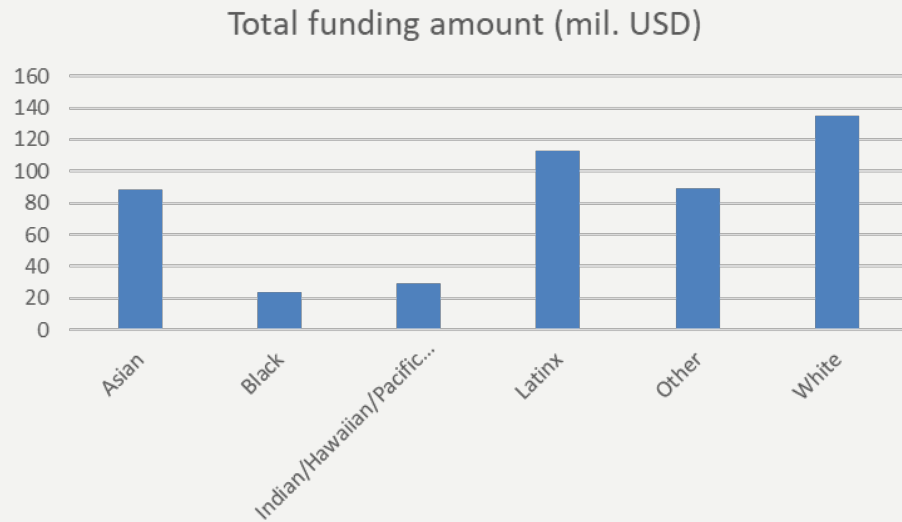
- The information collected by subgrantees varies, and there are differences in entrepreneur characteristics across subgrantees.
- Combining data from various sources would raise several problems in the interpretation of results:
 - ✓ Lack of representativeness (difficult to generalize the results)
 - ✓ Misunderstanding of information collected from different sources and environments
 - ✓ Difficulties in interpreting results
- Thus, for in depth-analysis for YR1, we focused on the [NEC community data merged to Crunchbase \(n=2,592\)](#) to answer questions about funding gaps by race/ethnicity.

Company Total Funding Amount



(Sample: NEC community & Crunchbase, n=2,592)

Company Total Funding Amount-early stage only



(Sample: NEC community & Crunchbase, only include Angel-Series A, n=299)