Constructing Sample Frames in Business Surveys When No Master Sample Frames Are Available

Jason Kosakow, Survey Director, Richmond Fed
Acacia Wyckoff, Research Analyst, Richmond Fed

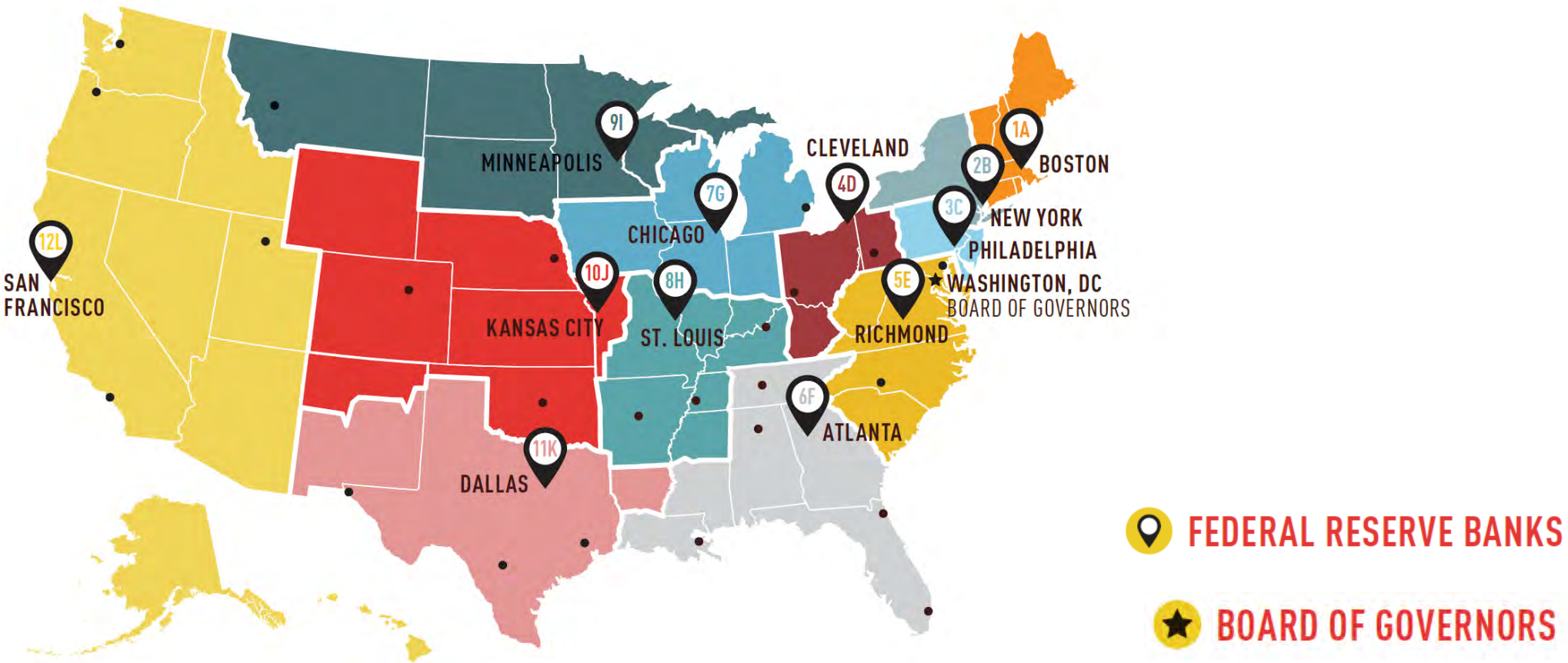
September 18th, 2024

The views and opinions expressed herein are those of the author. They do not represent an official position of the Federal Reserve Bank of Richmond or the Federal Reserve System.

FEDERAL RESERVE BANK OF RICHMOND
Richmond ▪ Baltimore ▪ Charlotte

Background Information

The United States Federal Reserve System



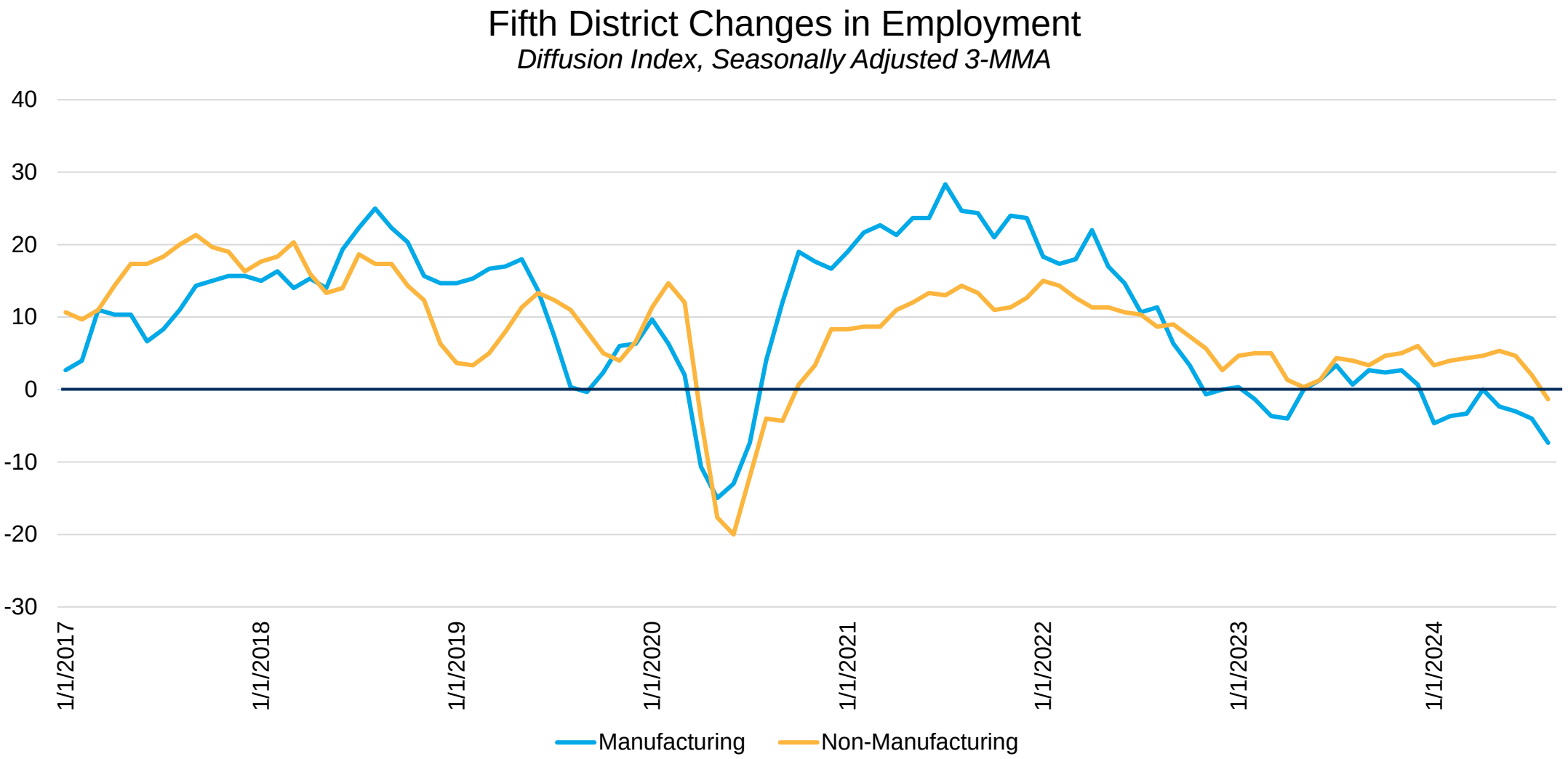
About the Fifth Federal District Business Surveys

- The Fifth Federal Reserve District covers the southern mid-Atlantic states
  - South Carolina, North Carolina, Virginia, most of West Virginia, D.C., Maryland
- Survey participation is **voluntary**
- We **cannot provide monetary incentives** for participation
- Mix of convenience and probability sample
- Recruitment is done to recruit new businesses into the panel
  - Business does not take survey at time of recruitment

These Surveys Are Important (At Least to Us)

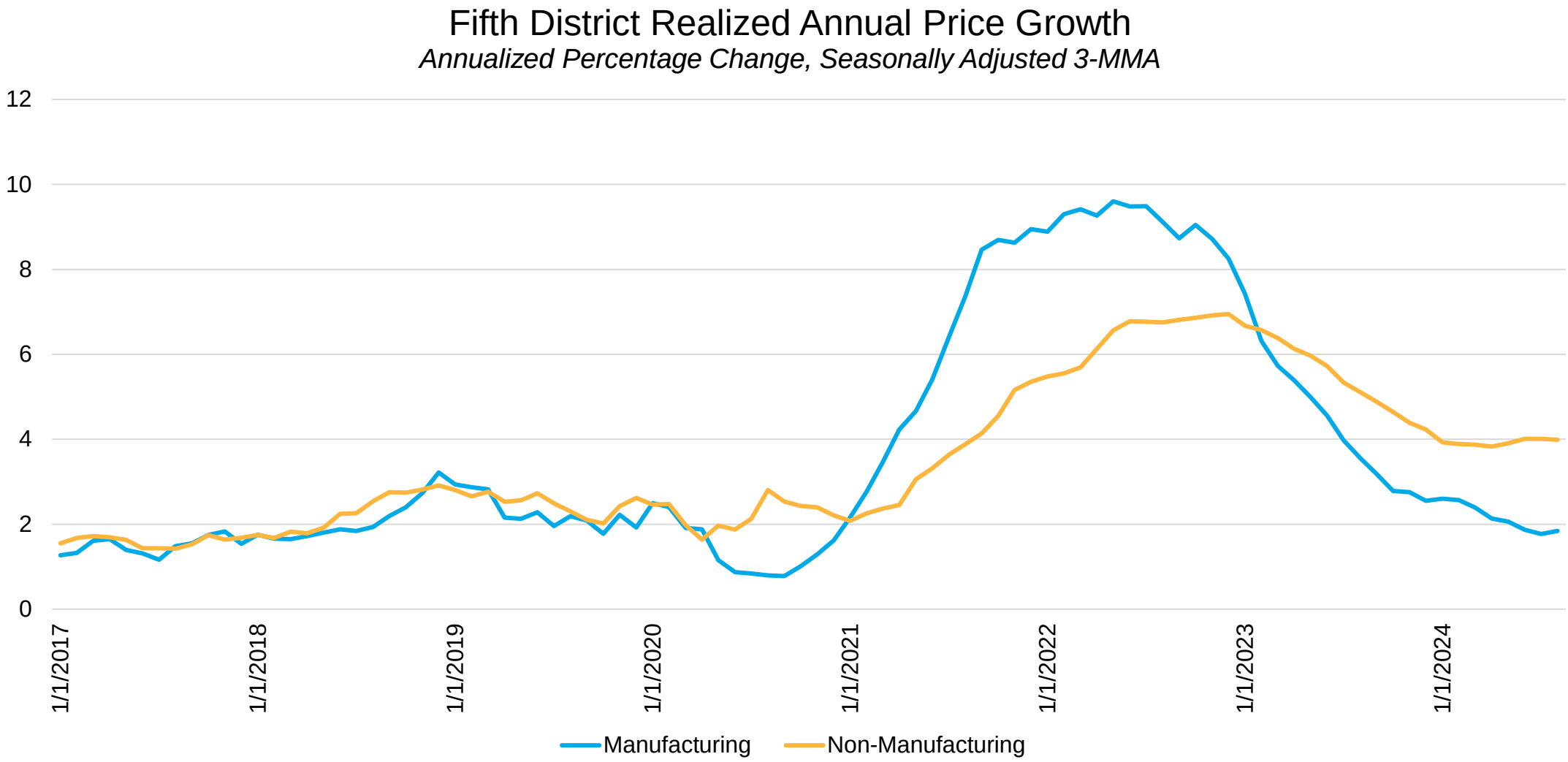
- **The FOMC Cycle is Every 6.5 weeks, so the Richmond Fed Needs Quick and Reliable Information.**
- **Richmond Fed Surveys Provide Real-Time Information About Business Conditions.**
  - Data provided by the government operates with a 1-month, 3-month, or yearly lag
- **Addition of "Special Questions" Specific to Relevant Economic Topics.**
  - Expected plans for capital expenditures, Frequency of changing prices, Labor availability
- **Information Collected is Used Across Multiple Publications.**
  - Press Releases
  - Beige Book
  - Articles
  - Economic Research

Real-Time Data to Inform Monetary Policy Decisions



Timeseries goes back to 1993

Real-Time Data to Inform Monetary Policy Decisions



FEDERAL RESERVE BANK OF RICHMOND
Richmond ▪ Baltimore ▪ Charlotte

Timeseries goes back to 1993

Coverage Issues When Administering Business Surveys

The Richmond Fed Relies on Multiple Modes of Business Recruitment

Richmond Fed Connections

- 1-on-1 Convos
- Industry Presentations
- Roundtables & Councils
- Partnerships

Traditional Recruitment Modes

- Cold Calling via Commercial Lists
- Cold Emails via Commercial Lists
- Mailings via Commercial Lists

FEDERAL RESERVE BANK OF RICHMOND
Richmond ▪ Baltimore ▪ Charlotte

Challenges When Using Commercial Vendor Lists

Challenge 1

High Rates of Inaccurate Contact Information

A frame with many out-of-scope units may result in a smaller number of completed interviews, which in turn may result in larger variances and costs per completed interview.

Challenge 2

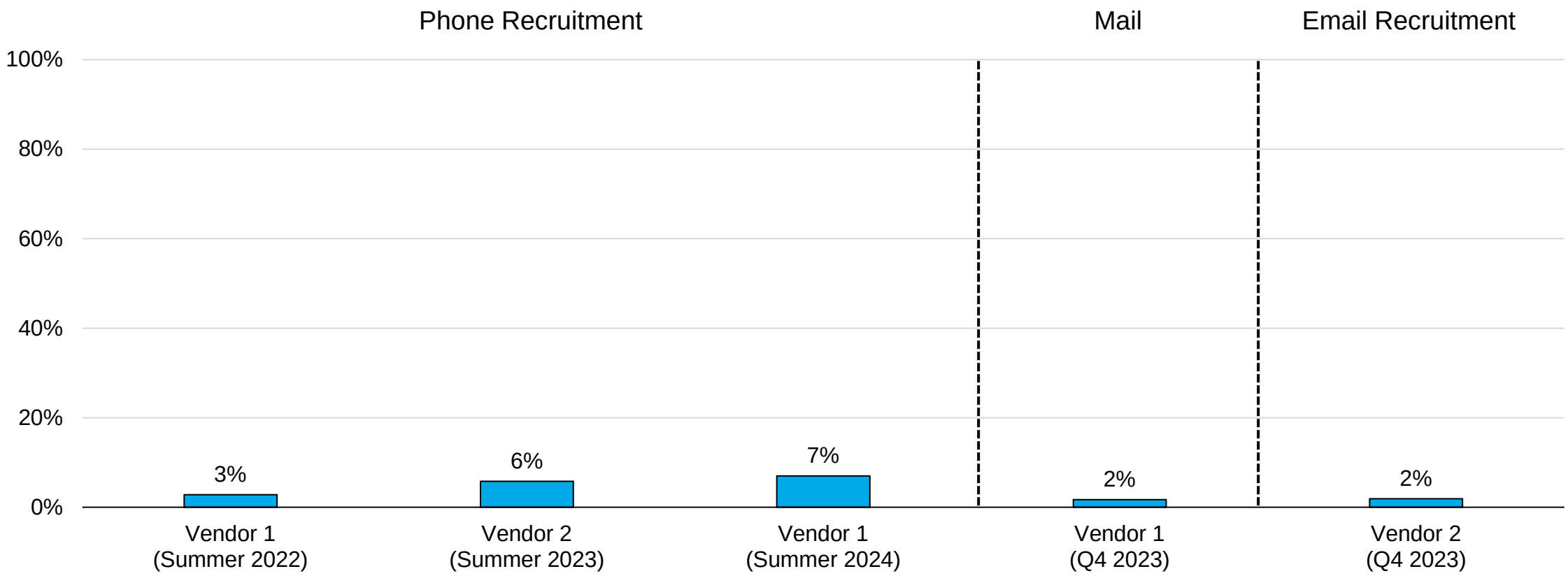
Unknown Converge Rates

Frames with systematic omissions may result in biased estimates.

AAPOR ABS Report, January 7, 2016

Like Household Surveys, Business Surveys Are Prone Have Low to Response Rates

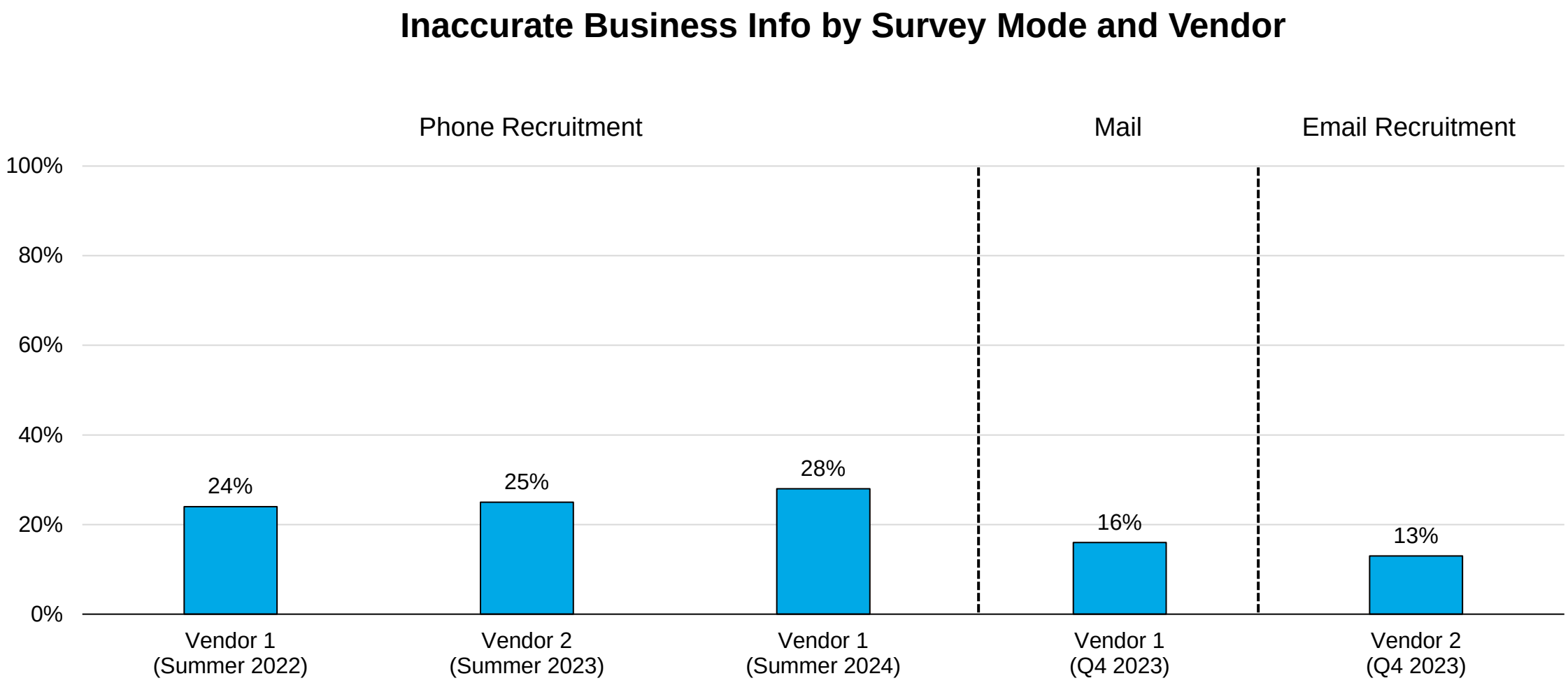
Recruitment Success by Survey Mode and Vendor
(Invalid Contacts Removed)



FEDERAL RESERVE BANK OF RICHMOND
Richmond ▪ Baltimore ▪ Charlotte

Calculation is total number of businesses recruited divided by total businesses reached out to

There Are High Rates of Inaccurate Business Info When Using Commercial Lists



FEDERAL RESERVE BANK OF RICHMOND
Richmond ▪ Baltimore ▪ Charlotte

Calculation is total number of businesses with inaccurate information divided by total businesses reached out to

Challenges When Using Commercial Vendor Lists

Challenge 1

High Rates of Inaccurate Contact Information

A frame with many out-of-scope units may result in a smaller number of completed interviews, which in turn may result in larger variances and costs per completed interview.

Challenge 2

Unknown Coverage Rates

Frames with systematic omissions may result in biased estimates.

AAPOR ABS Report, January 7, 2016

Important Questions Without Known Answers Regarding Unknown Coverage

Coverage Issues

?

Are there businesses that are systemically excluded from our sample frame?

?

Are the businesses that are excluded systematically different than those that are included?

Non-Response Issues

?

What effect does missing businesses in our sample frame have on our estimates?

How the Richmond Fed Attempted to Address Understanding Coverage Rates

Challenge

- At the moment, there are no publicly available business frames with known coverage.

Opportunity

- The Richmond Fed matched third-party commercial data to the USPS CDSF.

Outcome

- A sample frame with known levels of coverage.
- The assumption is the USPS address file has high-levels of coverage.

This Methodology is Common for Household Surveys, but Has Only Been Tested Once (O'Brien 2013, EIA) for Establishment Surveys

Constructing the Frame

Step 1: Selecting Geographies

- We purchased the USPS CDSF (filtered for businesses) for three geographies
  - Washington, D.C. (city only)
  - Asheville, NC MSA
  - Greenville, SC MSA
- We needed diverse geographies within our District to get a sense if this methodology would be appropriate.
- These three geographies offer diversity in:
  - Urban versus Rural
  - Industry composition
  - Population/Demographics

Step 2: Finding Data Sources to Append to the USPS File

The USPS list does not contain firmographic information, so we matched third-party data to addresses in the USPS file.

Data Source	Sectors Covered	Availability
Vendor 1	All	Via a License
Vendor 2	Manufacturing	Via a License
Vendor 3	All	Via a License
IRS exempt organization business master file	Tax exempt organizations (usually nonprofits, religious organizations, etc.)	Public
YellowPages.com	Whatever is listed	Webscraping

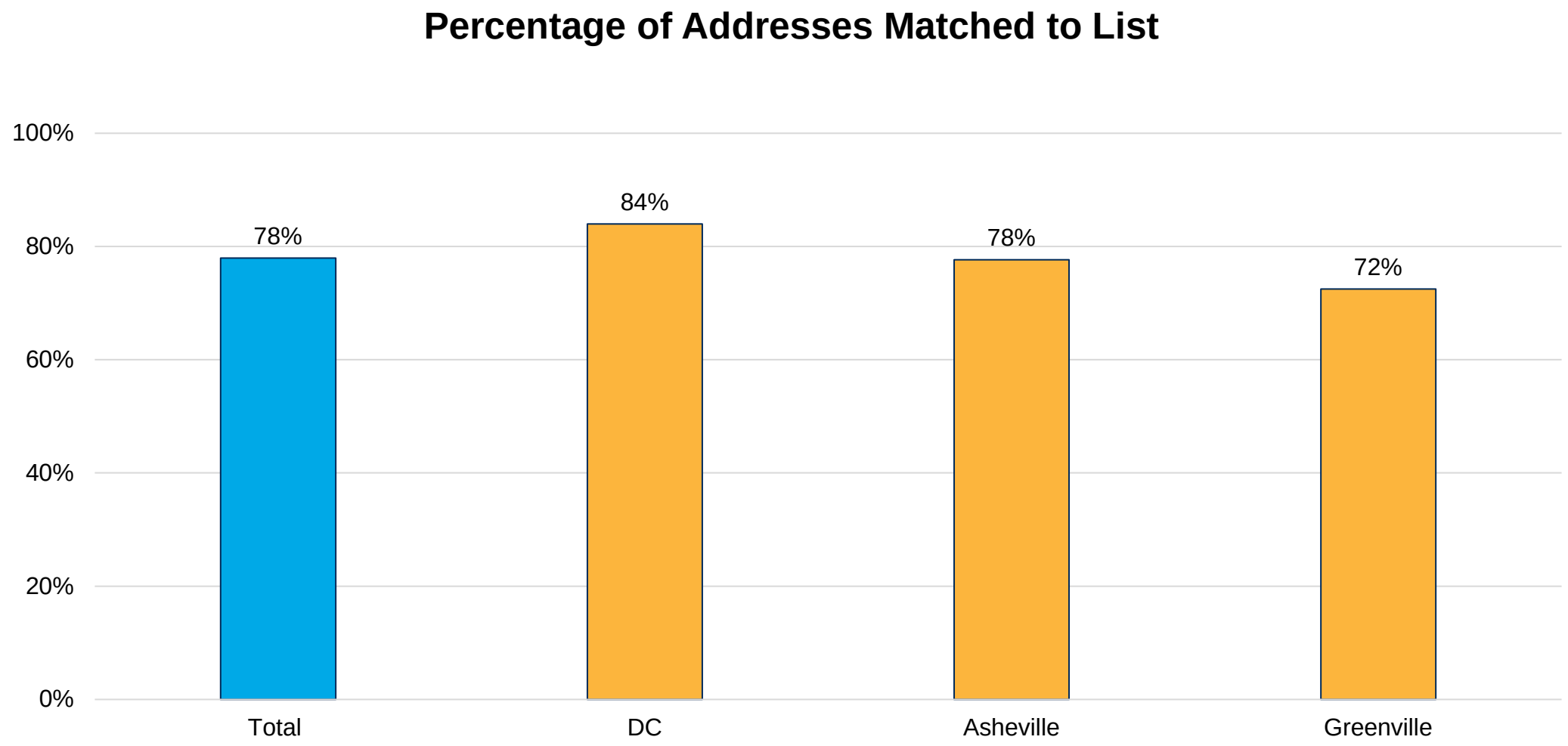
Data Availability for Each Data Source

Sample Frame Source	Business Address	Business Name	Contact Name	Phone Number	Email	Industry	Other Demos
Vendor 1							
Vendor 2							
Vendor 3							
IRS exempt organization business master file							
YellowPages.com							

Available | Available For Some | Not Available

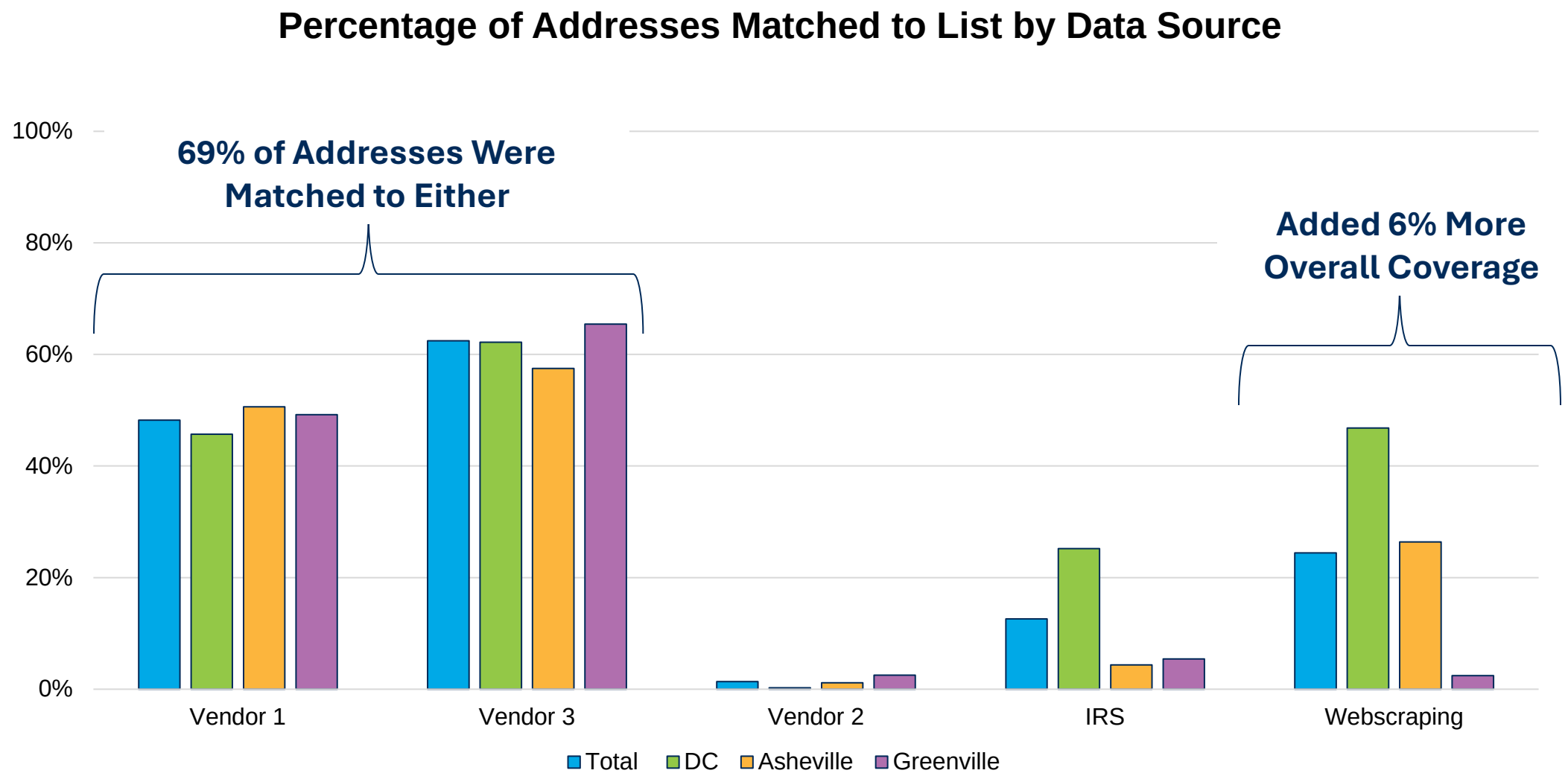
Findings

We Were Able to Match Nearly 80 Percent of Addresses

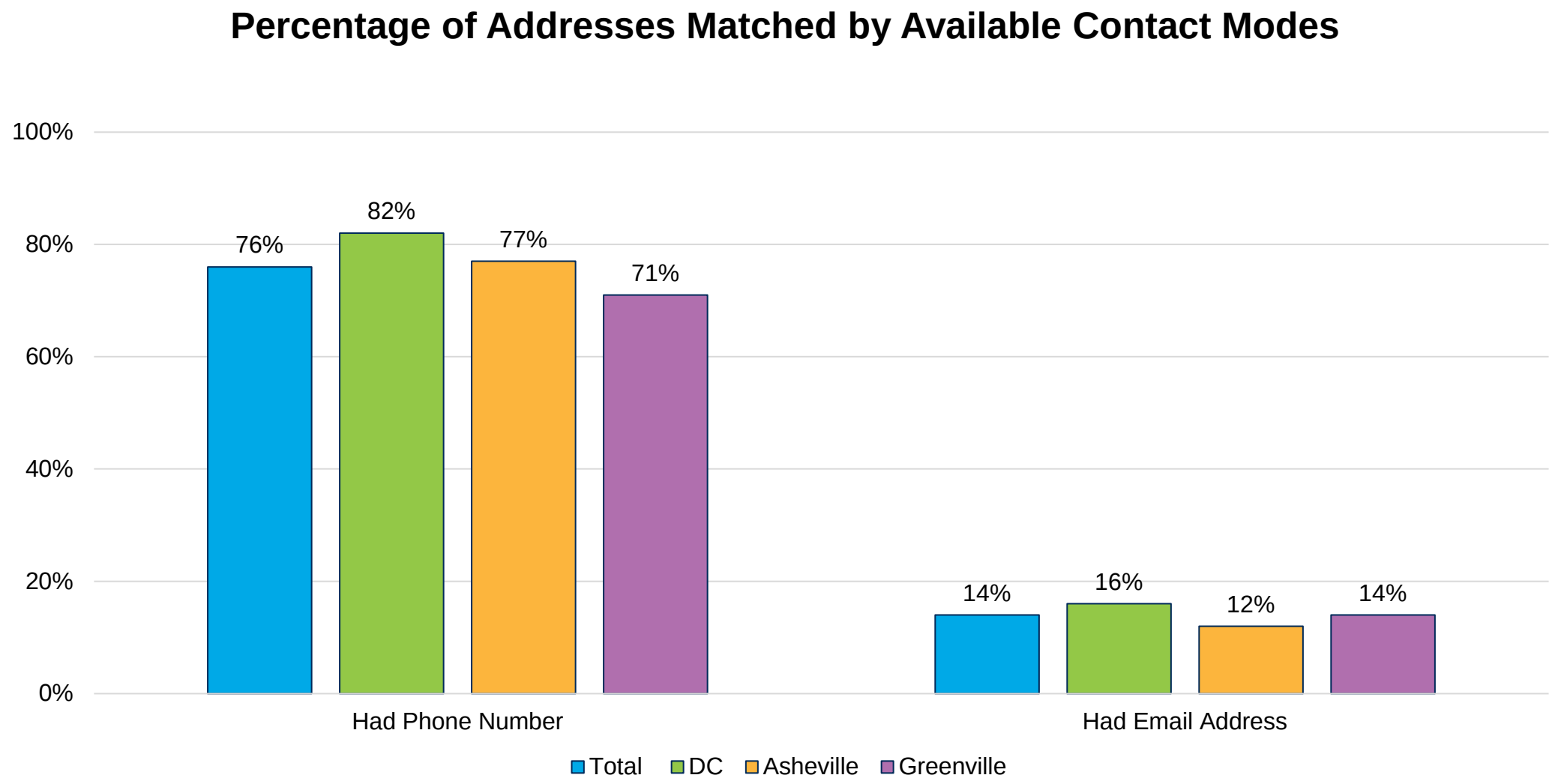


FEDERAL RESERVE BANK OF RICHMOND
Richmond ▪ Baltimore ▪ Charlotte

Vendor 3 Had the Highest Match Rates Among All Data Sources

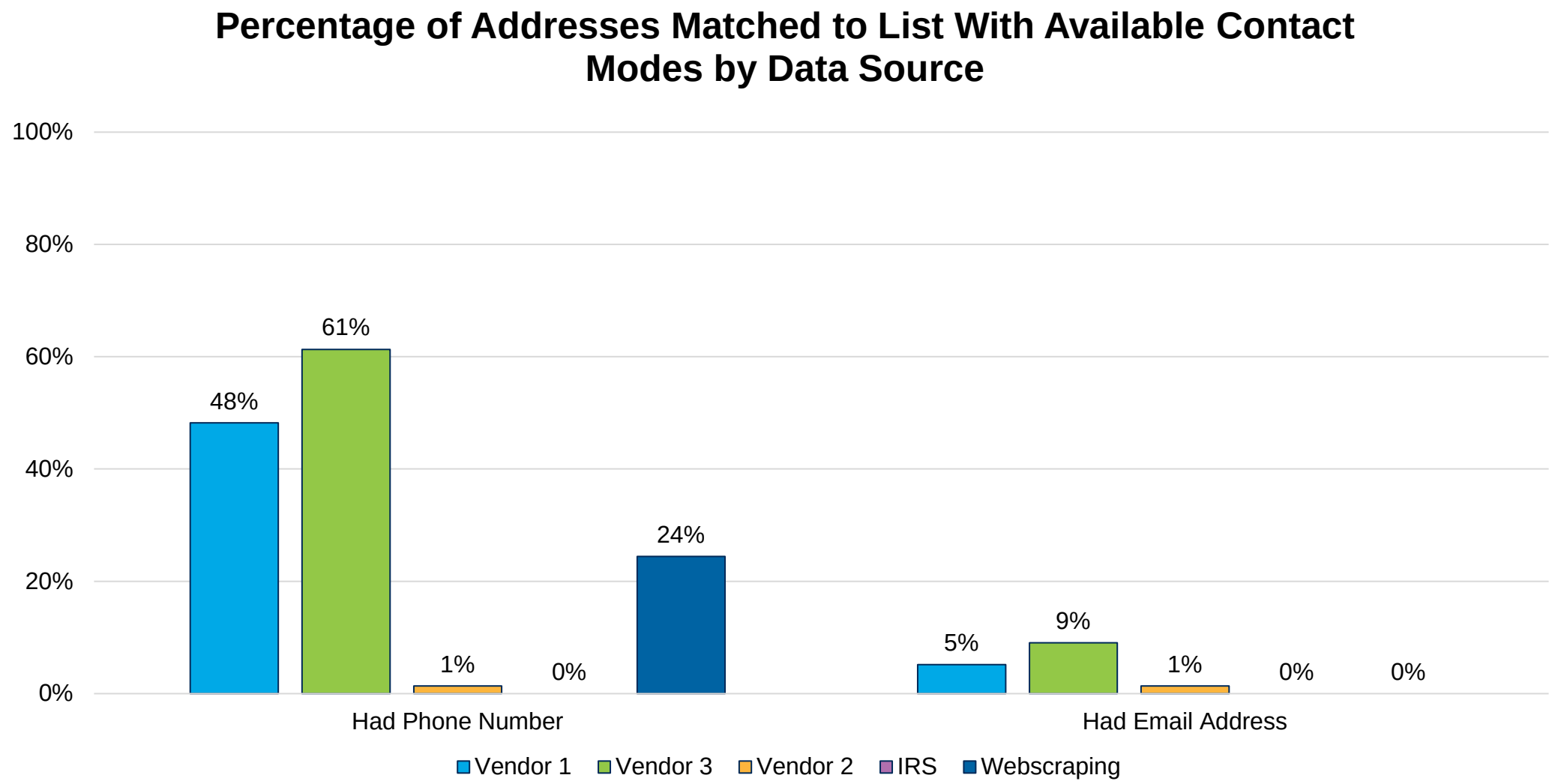


Greenville Had Lower Phone Number Matches But Similar Email Match



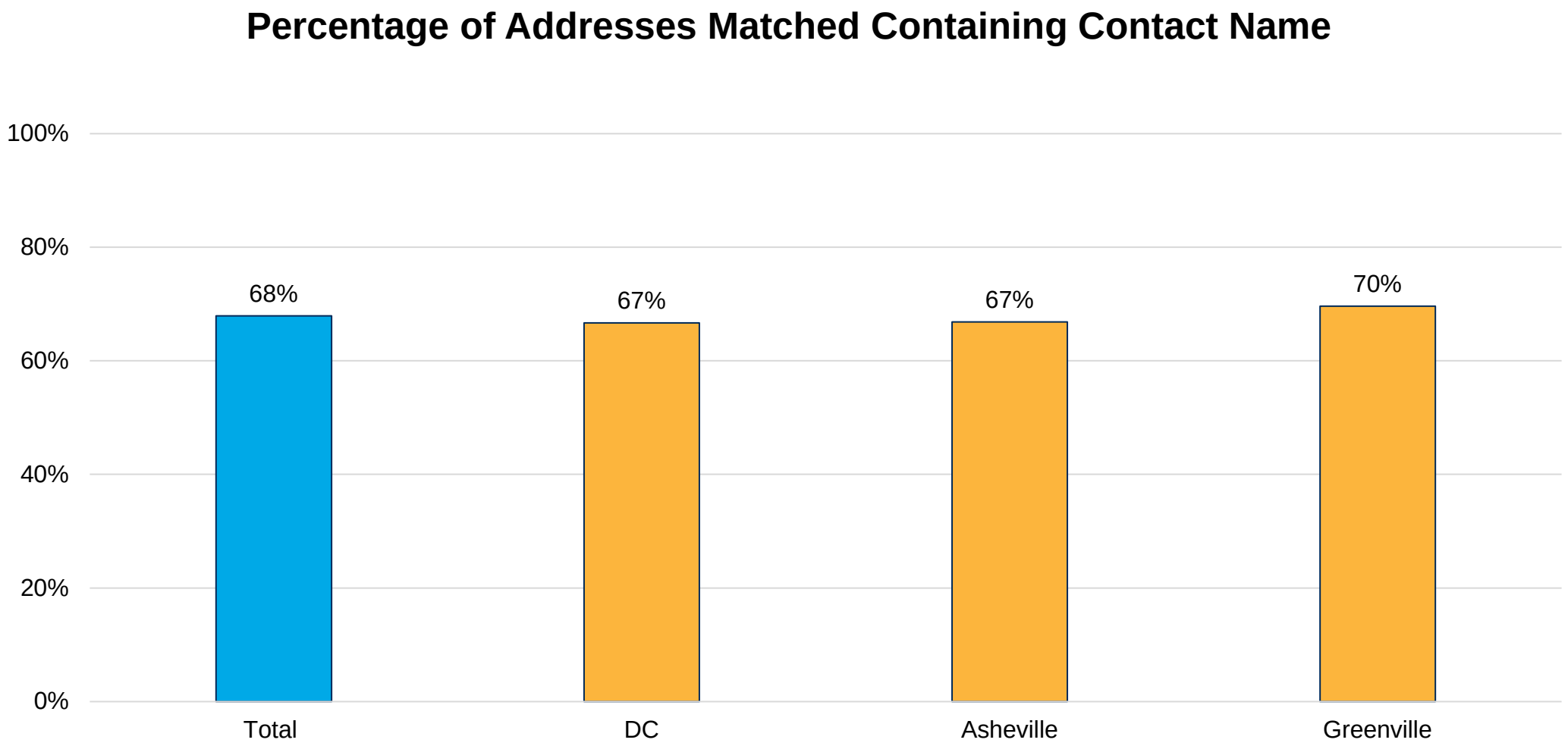
FEDERAL RESERVE BANK OF RICHMOND
Richmond ▪ Baltimore ▪ Charlotte

Contact Information Was Most Available From Vendor 3



FEDERAL RESERVE BANK OF RICHMOND
Richmond ▪ Baltimore ▪ Charlotte

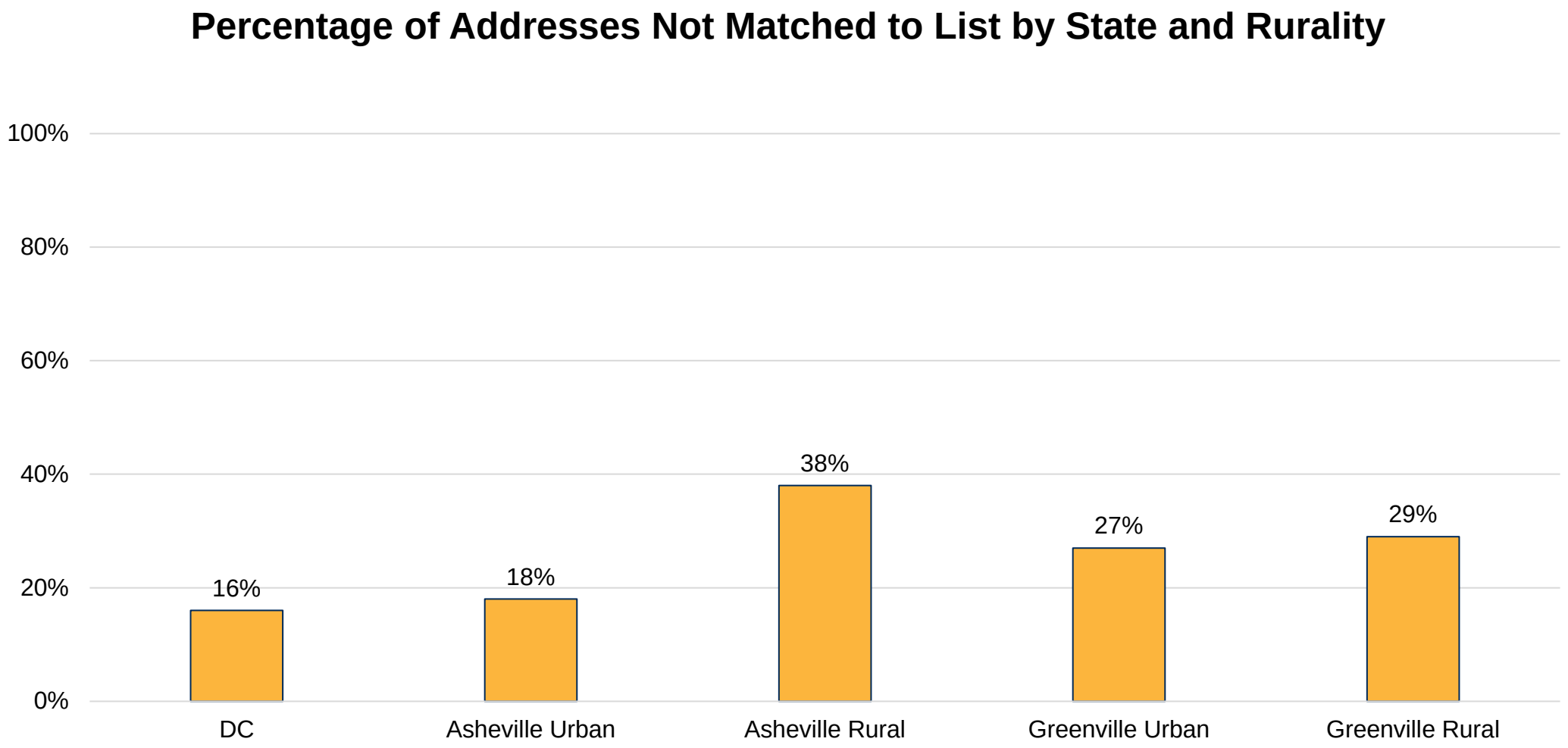
Contact Name Availability Was Similar Across Geographies



FEDERAL RESERVE BANK OF RICHMOND
Richmond ▪ Baltimore ▪ Charlotte

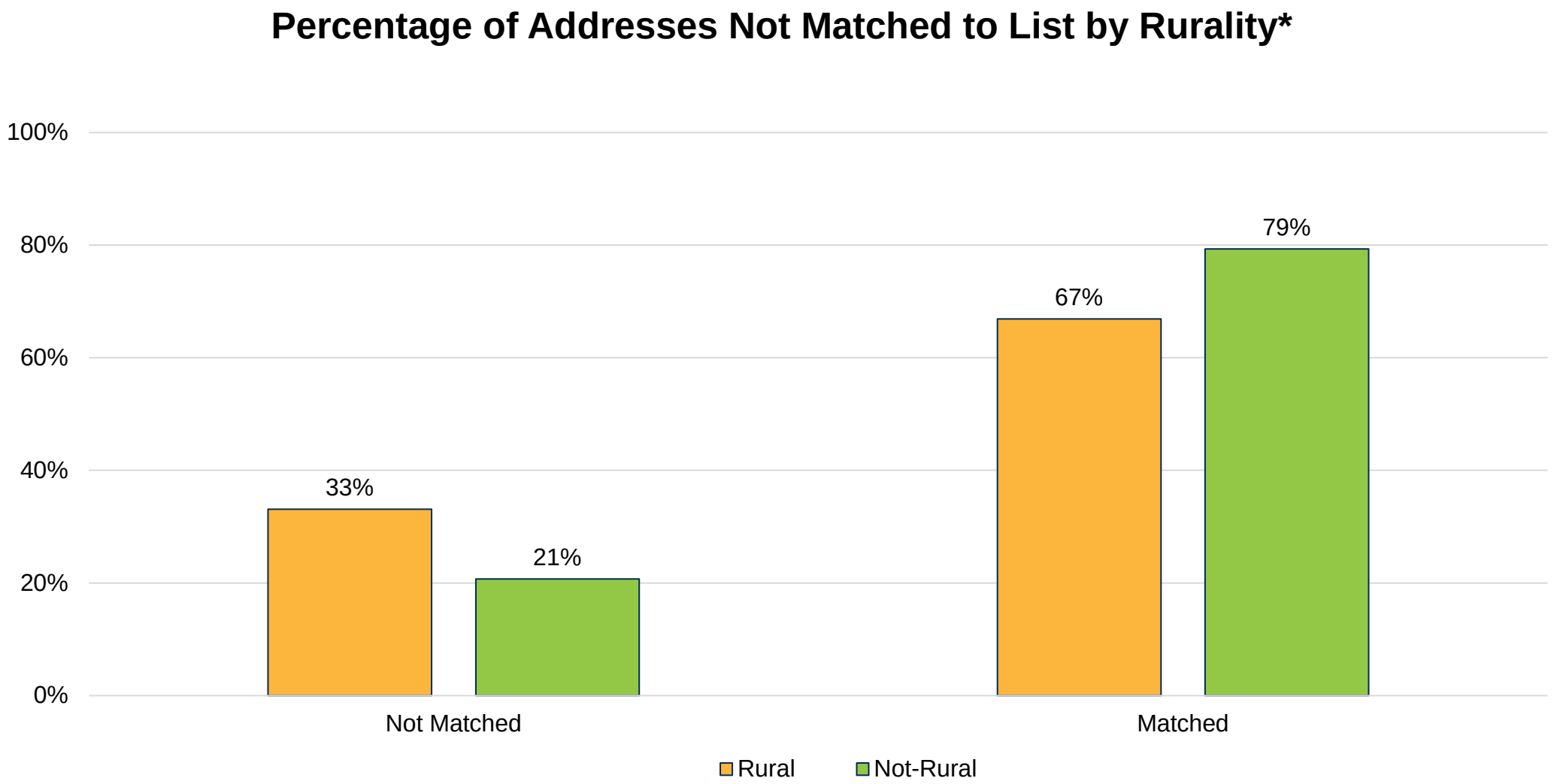
Information on Non-Matched Addresses

Rural Asheville Had the Highest Non-Match Rate



FEDERAL RESERVE BANK OF RICHMOND
Richmond ▪ Baltimore ▪ Charlotte

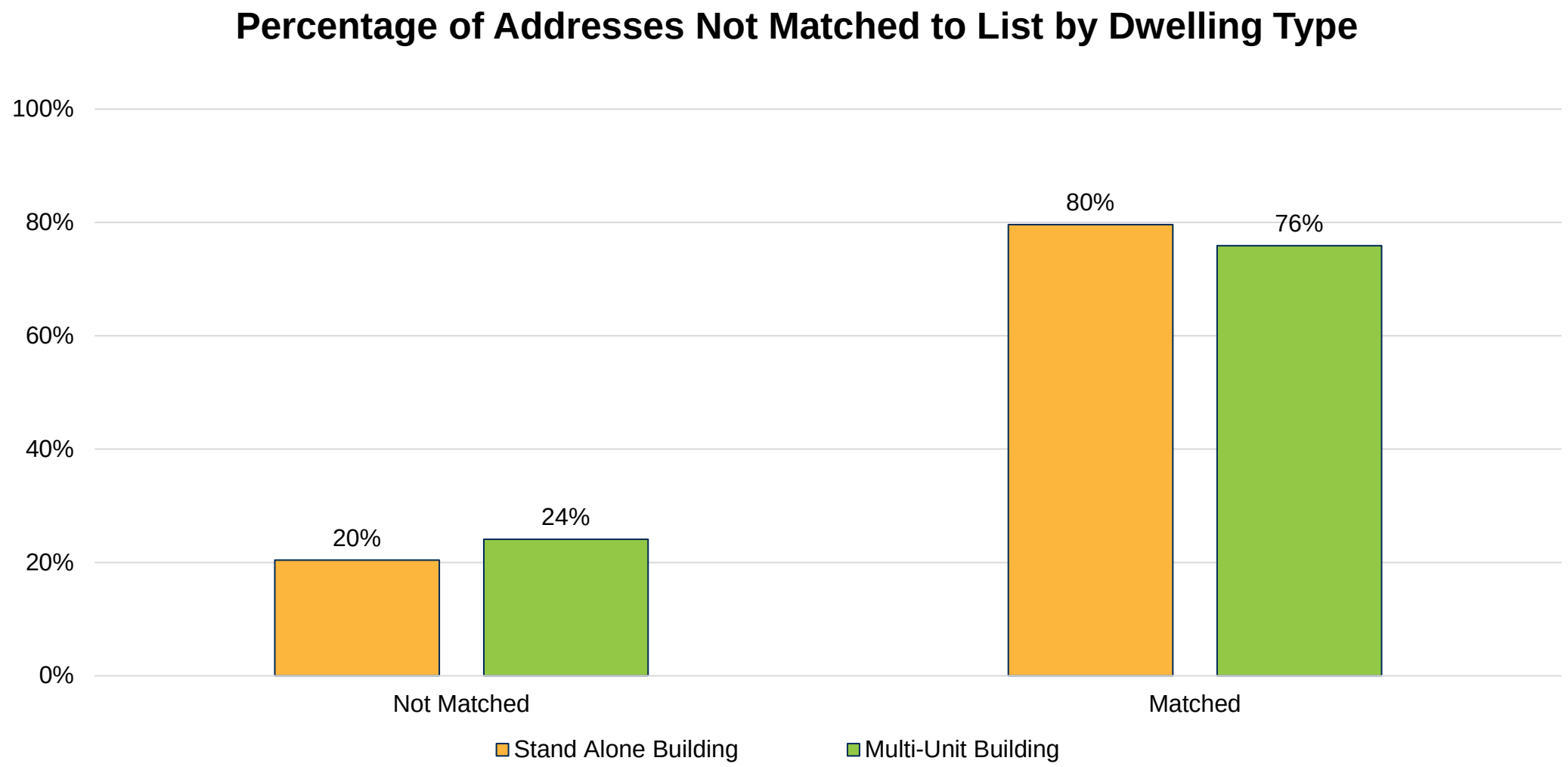
Rural Firms Were More Likely to Not Be Matched



** Rural defined as a U.S. Census-designated urban areas containing 25,000 or fewer people*

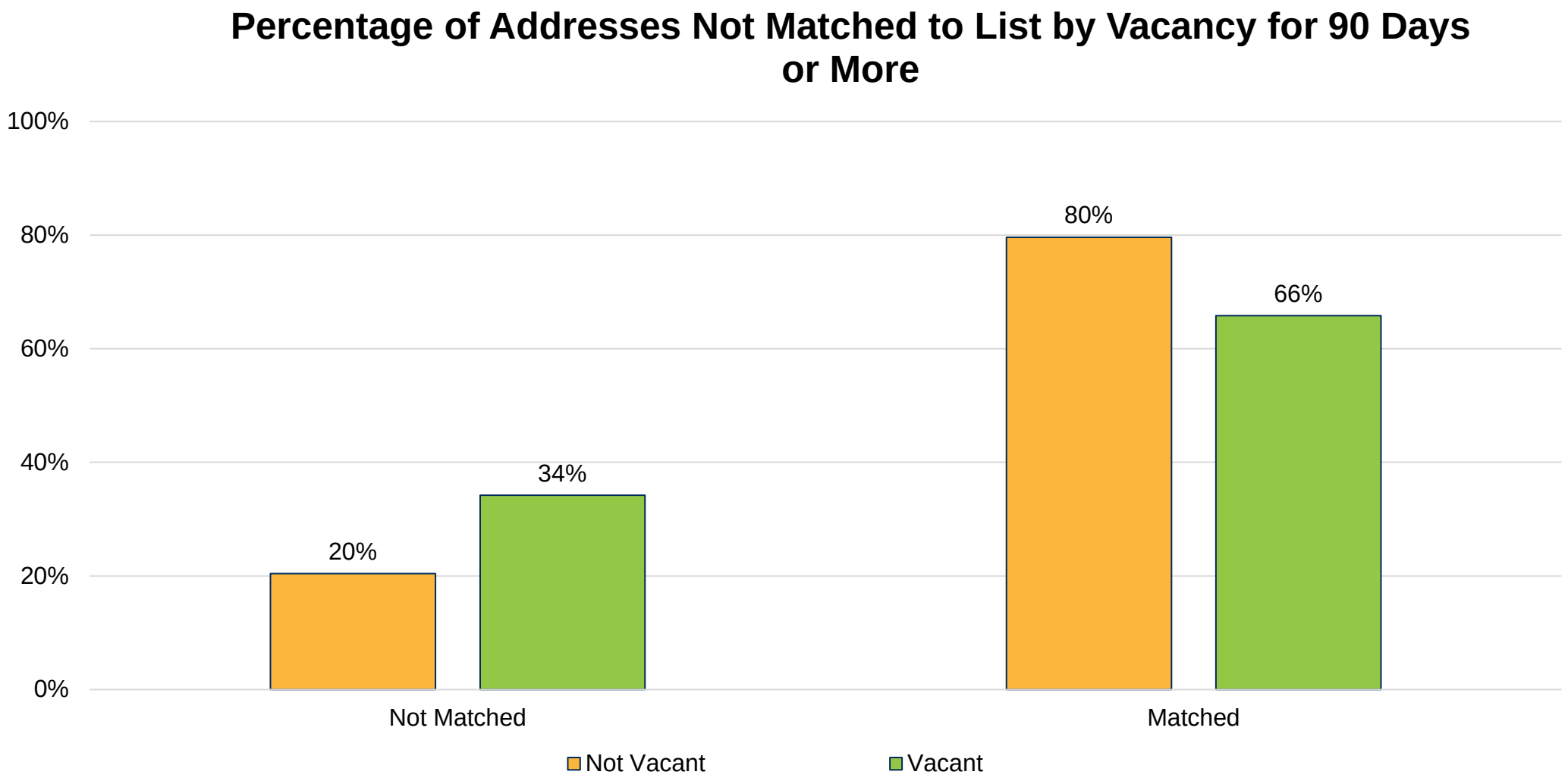
FEDERAL RESERVE BANK OF RICHMOND
Richmond ▪ Baltimore ▪ Charlotte

Multi-Unit Buildings Were Less Likely to Be Matched



FEDERAL RESERVE BANK OF RICHMOND
Richmond ▪ Baltimore ▪ Charlotte

Vacant Buildings Less Likely to Be Matched



FEDERAL RESERVE BANK OF RICHMOND
Richmond ▪ Baltimore ▪ Charlotte

Third-Party Data Matching & Quality

Two Size Categories Show Differences Against the CBP Benchmark

	Washington, D.C.		Asheville, NC MSA		Greenville, SC MSA	
Establishment Size	**USPS List**	**2022 CBP**	**USPS List**	**2022 CBP**	**USPS List**	**2022 CBP**
Fewer Than 5 Employees	50.5%	50.9%	57.4%	57.8%	55.4%	53.8%
5 to 9 Employees	21.1%	17.0%	22.1%	17.4%	22.3%	17.8%
10 to 19 Employees	14.3%	13.3%	11.4%	12.0%	12.2%	13.0%
20 to 99 Employees	11.7%	15.1%	7.7%	11.0%	8.6%	12.8%
100 to 499 Employees	2.2%	3.4%	1.3%	1.6%	1.4%	2.2%
500 or More Employees	0.3%	0.4%	0.2%	0.2%	0.1%	0.5%
Total	16,559	23,874	10,294	13,131	18,827	22,457
Emp Size Missing *(Excluded from Above)*	34.7%	-	32.3%	-	30.8%	-

Shaded areas are 3 percentage points or more different than CBP

Distribution of Matched Businesses Against Benchmark

Industry (Washington, D.C.)	USPS List	2022 CBP
Agriculture, Forestry, Fishing and Hunting	0.2%	0.0%
Mining, Quarrying, and Oil and Gas Extraction	0.1%	0.0%
Utilities	0.3%	0.3%
Construction	5.1%	2.2%
Manufacturing	3.0%	0.4%
Wholesale Trade	2.4%	1.6%
Retail Trade	10.8%	6.9%
Transportation and Warehousing	1.6%	0.7%
Information	3.7%	3.6%
Finance and Insurance	5.5%	4.8%
Real Estate and Rental and Leasing	5.1%	6.3%
Professional, Scientific, and Technical Services	16.1%	23.9%
Management of Companies and Enterprises	0.1%	0.9%
Administrative and Support and Waste Management and Remediation Services	5.0%	4.5%
Educational Services	1.8%	2.8%
Health Care and Social Assistance	9.0%	9.8%
Arts, Entertainment, and Recreation	1.3%	1.9%
Accommodation and Food Services	6.2%	11.6%
Other Services (except Public Administration)	21.2%	17.9%
Industries not classified	1.7%	0.1%
Total	18,859	23,873
Industry Missing	25.6%	-

Shaded areas are 3 percentage points or more different than CBP

Distribution of Matched Businesses Against Benchmark

Industry (Asheville)	USPS List	2022 CBP
Agriculture, Forestry, Fishing and Hunting	0.8%	0.1%
Mining, Quarrying, and Oil and Gas Extraction	0.2%	0.0%
Utilities	0.2%	0.2%
Construction	5.7%	11.3%
Manufacturing	4.7%	3.7%
Wholesale Trade	4.6%	3.7%
Retail Trade	18.7%	13.6%
Transportation and Warehousing	2.1%	1.8%
Information	1.5%	1.6%
Finance and Insurance	5.9%	4.4%
Real Estate and Rental and Leasing	4.8%	7.2%
Professional, Scientific, and Technical Services	9.8%	11.2%
Management of Companies and Enterprises	0.2%	0.3%
Administrative and Support and Waste Management and Remediation Services	4.3%	5.6%
Educational Services	1.8%	1.8%
Health Care and Social Assistance	10.0%	12.6%
Arts, Entertainment, and Recreation	1.8%	2.0%
Accommodation and Food Services	7.7%	9.6%
Other Services (except Public Administration)	13.5%	9.3%
Industries not classified	1.8%	0.1%
Total	10,630	13,131
Industry Missing	30.1%	-

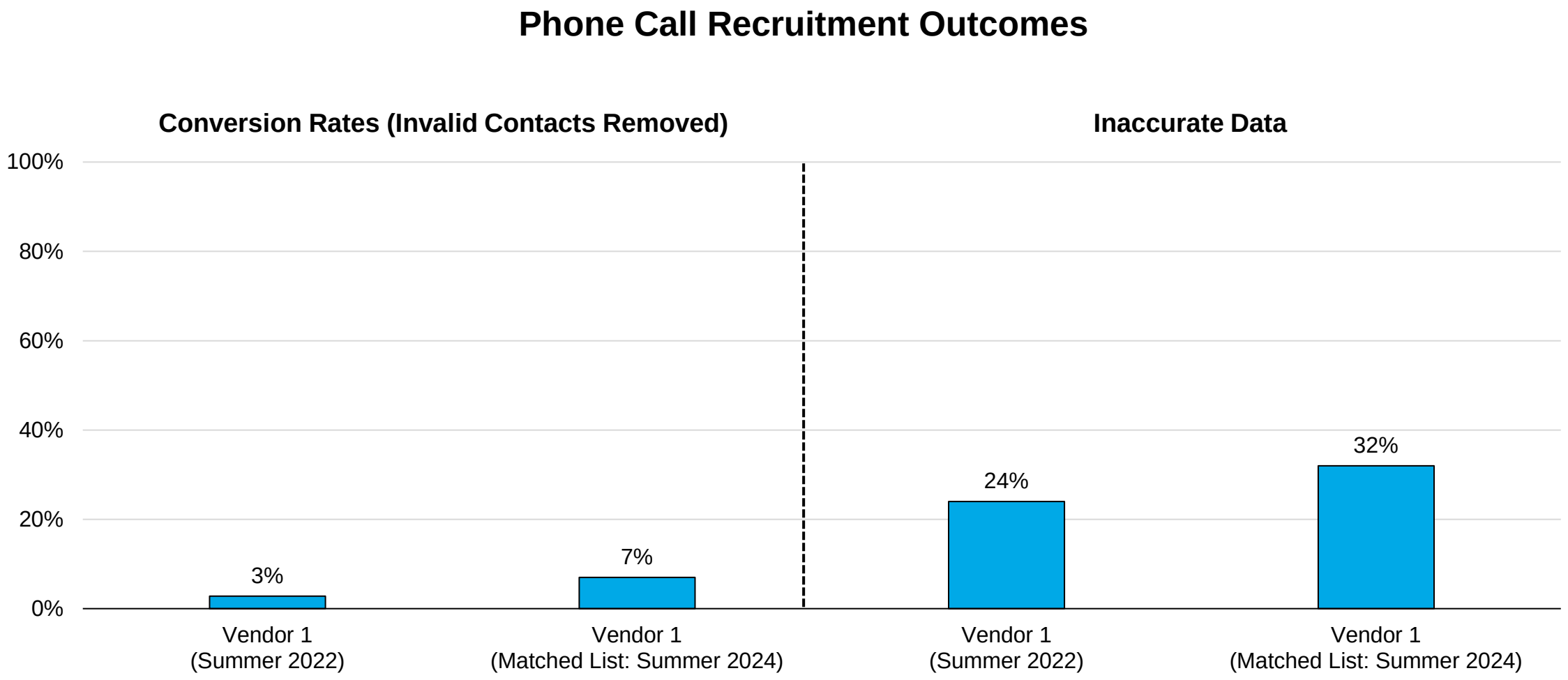
Shaded areas are 3 percentage points or more different than CBP

Distribution of Matched Businesses Against Benchmark

Industry (Greenville)	USPS List	2022 CBP
Agriculture, Forestry, Fishing and Hunting	0.3%	0.2%
Mining, Quarrying, and Oil and Gas Extraction	0.1%	0.1%
Utilities	0.2%	0.2%
Construction	5.0%	9.7%
Manufacturing	4.3%	4.1%
Wholesale Trade	3.7%	4.7%
Retail Trade	14.3%	13.9%
Transportation and Warehousing	1.8%	2.5%
Information	0.7%	1.4%
Finance and Insurance	4.8%	6.6%
Real Estate and Rental and Leasing	2.6%	5.9%
Professional, Scientific, and Technical Services	5.5%	11.4%
Management of Companies and Enterprises	0.1%	0.6%
Administrative and Support and Waste Management and Remediation Services	5.8%	5.7%
Educational Services	2.8%	1.3%
Health Care and Social Assistance	13.3%	9.8%
Arts, Entertainment, and Recreation	2.5%	1.5%
Accommodation and Food Services	10.1%	9.7%
Other Services (except Public Administration)	20.1%	10.5%
Industries not classified	2.0%	0.1%
Total	19,507	22457
Industry Missing	28.3%	-

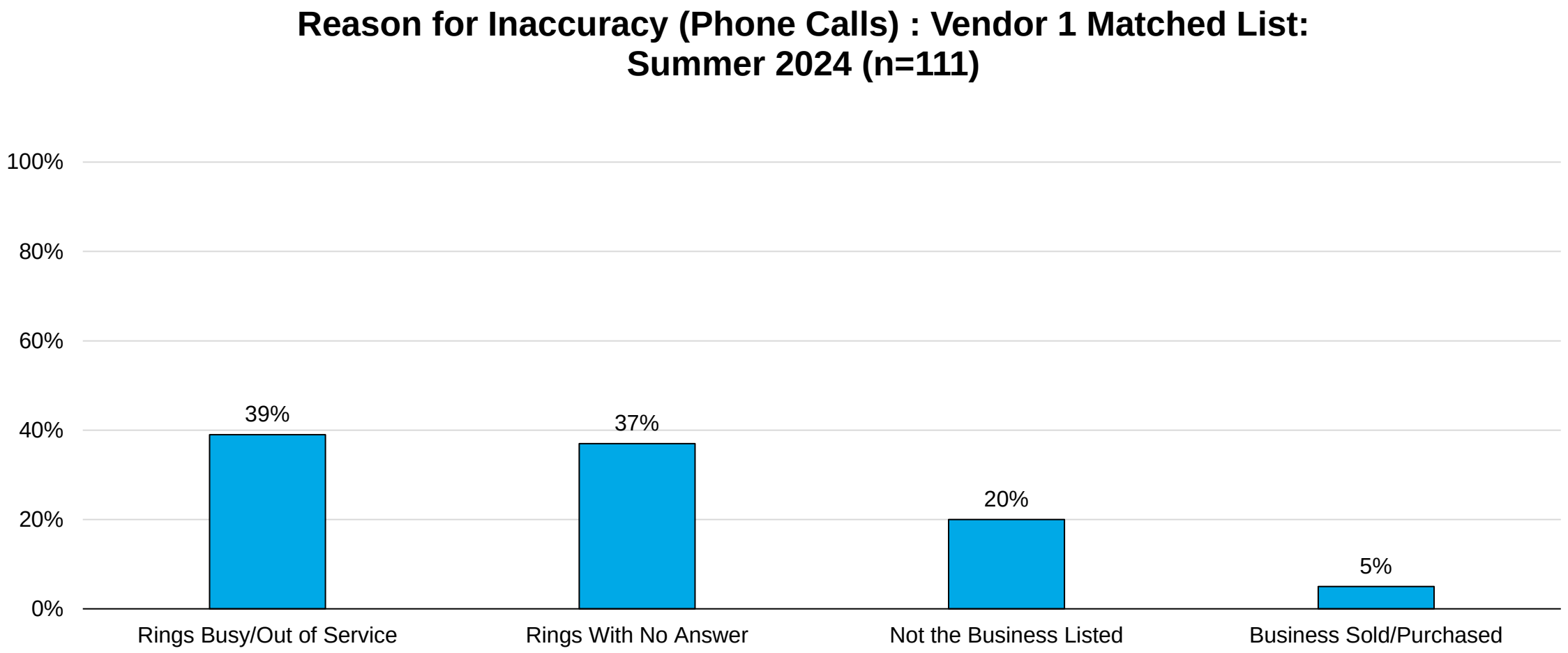
Shaded areas are 3 percentage points or more different than CBP

Known Coverage Does Not Mean Better Business Accuracy



Vendor 1 Matched List: Summer 2024: n=352

Most Inaccurate Business Information Is Due to Phone Numbers Issues



FEDERAL RESERVE BANK OF RICHMOND
Richmond ▪ Baltimore ▪ Charlotte

Concluding Remarks & Next Steps

Conclusions

- **The USPS CDSF could be a viable option for frame construction**
- **This methodology does provide a "known" coverage rate**
  - Almost 4-in-5 addresses were matched with third-party data
  - Like other ABS studies, certain subgroups had lower match rates
    - Rural businesses, Businesses in multi-unit dwellings
  - More work needs to be done on matching the list with firmographic data (1/3 of firms are not matched with an industry or firm size)
- **Data quality will continue to be an issue**
  - Data quality for commercial business lists has not been explored publicly
    - This creates difficulties in knowing which data to merge with the USPS list
- **Webscraping can be used to increase coverage in targeted geographies**
  - Scraping the entire internet is not feasible, so selecting low-coverage areas and locating public business information could increase coverage rates

Data Merging Issues

- Merging USPS data with third party data suffers from similar issues as household matching
  - Multiple establishments with the same address (some examples were shopping malls, hospitals, or buildings with many different suites and offices)
    - 30% of addresses have multiple businesses listed; 7% of addresses have 3 or more businesses listed
    - For example, not all Vendor 3 data has office numbers listed for DC Children's Hospital, so there are 807 doctors at the same overarching address
  - Different sources have different ways of labeling floors, offices, suites, etc.
    - Basic corrections done to remove special characters, move all addresses to upper case so nothing is case sensitive
  - For now, we have cut the data to include only one instance of each address with one business per source
    - We have an accurate list of WHICH addresses in USPS list have a match, further work needs to be done to determine HOW MANY matches per address

Please Reach Out!

Jason.Kosakow@rich.frb.org **Acacia.Wyckoff@rich.frb.org**

FEDERAL RESERVE BANK OF RICHMOND
Richmond ▪ Baltimore ▪ Charlotte