

# Module 3



## **GENERATING AND USING PRIMARY DATA**

# Augmenting Data with Local Efforts



- Lots of secondary sources, but where does local data add value?



# Overview of Module 3



- Guiding primary data collection after considering your local food assessment needs and mission
- Sampling techniques (i.e., who will receive the questions you pose), including what each approach entails and the associated pros and cons
- Qualitative and quantitative data collection methods, including a detailed description of data collection methods, tips to collect unbiased and valuable data and uses for each approach
- Preliminary data analysis techniques

# Before you Begin Collecting Data



- Defined scope, specific goals, objectives, timeframe, resources & regional boundaries (module 1)
- Used secondary data to understand baseline conditions or potential economic contribution
  - Found that you still need additional data to accomplish the goals of your assessment
- Examined available secondary data
  - Determined that data does not exist for the specific question(s) you seek to answer, or
  - Available secondary data does not reflect conditions in the study area

# Words of caution before it begins...



- Team determined it has substantial time, resources, and expertise to devote to collecting and analyzing primary data.
- Realized that primary data collection, analysis and interpretation requires skill and training
- Understands that it often costs, at minimum, several thousand dollars to conduct even a small study
  - You may need to hire people to test and administer the surveys/interviews, pay for travel, compensate respondents for their time, etc.

# Dimensions and Indicators



- **Indicators of Economic Prosperity**

- Contributes to vendor viability;
- Purchases product from other neighboring businesses;
- Improves food affordability

- **Indicators of Social Interaction**

- Diversity (race, age, gender, ethnicity, culture) of vendors and shoppers;
- Provides a community gathering space;
- Has an open and democratic governance structure

# Dimensions and Indicators



- **Indicators of Public Health**
  - Provides access to healthful foods;
  - Provides cooking lessons and other nutrition education opportunities;
  - Encourages biking and walking to farmers' markets
- **Indicators of Environment and Aesthetics**
  - Provides green space;
  - Is visually attractive;
  - Sells environmentally sound & organically grown products;
  - Educates the public and models sustainable practices such as recycling and composting at farmers' markets



# Variables and Samples



- Census of all relevant people/enterprises is challenging except among very small niches
  - Instead you need data on variables from a sample
- Representative samples-what number and type of responses will be generalizable?
  - Value of a random sample is the ability to credibly generalize results to the overall population
- Probabilistic approach gathering information from only a share of possible respondents
  - May be targeted to those of a certain type, demographic, or behavior



# Non-Representative Sample



- **Key Informants** are those who have substantial knowledge about your subject of interest.
  - Vendors at a farmers' market, farmers selling to farm-to-school programs or elected officials of your county or state.
- One commonly used strategy for selecting a key informant sample is a technique to ensure that the sample represents the diversity of the population along several dimensions.
  - For example, if you are studying firms, you might look at **firmographic** dimensions like geographic location, size of business, years in business, products sold, and number of employees.

# Snowball Sampling



- Once you identify someone who is knowledgeable on a subject, it can be helpful to ask the subject for suggestions of other people who are knowledgeable, and have meaningful experiences or perspectives
  - *Snowball sampling* often goes hand in hand with key informant sampling.
- For example, when learning about a specialized product, you may ask your informant about the people he/she buys product from or sells to.
  - The size of the sample grows over time as new contacts are identified by informants, just as a snowball gains size and momentum when rolling down a hill

# Quota Sampling



- ***Quota sampling ensures*** that the sample resembles the population by establishing quotas or minimum thresholds for specific characteristics.
  - For example, if the population of farmers' market vendors in your county is 50% male and 50% female, divided between 75% farmers and 25% vendors of prepared food, you may recruit subjects for your survey until your sample meets these criteria.
- While this sample will not be representative, this method will still enable you to avoid some bias
  - This sampling technique may also be used to oversample minority populations or seek out divergent viewpoints to ensure you hear a broad array of viewpoints.

# Convenience Sampling



- **Convenience sampling** selects subjects who are easily accessible.
  - The purpose is to obtain a sample with minimal cost and effort expended.
- This is a very common method of sampling and encompasses many possible strategies
  - It may involve sampling shoppers at a farmers' market on certain days by setting up a booth or standing with a clipboard and asking for participation.
  - It may involve hanging up signs at the local health food stores, or sending an email with a link to an on-line survey to a group of farmers, vendors, members, or other population of interest.

# Other approaches to gain validity



- Some researchers choose to administer **intercept surveys** at markets on a variety of days and times
- You may also want to adopt the practice of **triangulating** your findings – i.e., asking the questions from different sources
  - For any specific response to be a **significant observation**, the response must be raised by several respondents independently
  - While this technique will not give you a random sample, it may well give you excellent insights about key issues
- These opinions may not always be correct, but they are likely to be valuable, place-specific insight

# Data Collection



- **Qualitative** research collects data detailing the quality of someone's experiences, very often in the form of interviews.
  - Qualitative data deals with descriptions and data that cannot be measured using numbers.
- **Quantitative** research deals with numbers and data that can be measured.
  - commonly use surveys in which subjects choose answers best corresponding to their experience.
  - Often statistical analysis can be performed comparing frequencies and finding relationships among responses.

# Qualitative Methods



- **Observations and interviews.**
  - Observations involve going to an unfamiliar setting and then watching and observing what unfolds.
- **There is a broad array of things to look for, such as:**
  - Participants (who is there, how many, what are their demographic attributes);
  - Behaviors (what do they do, for how long);
  - Interactions (who do they talk to, work with, what is the non-verbal communication happening);
  - Physical environment (sights, sounds, climate, location);
  - Outcomes (what happens as a result).



# Qualitative Interviews



- Interviews may take the form of one-on-one discussions, or having a focus group
  - The purpose of conducting interviews is to discover the interconnection of ideas and behaviors
- It is good to assemble a formal questionnaire but you may also ask additional probing questions
  - Get deeper information on certain themes
  - Open ended questions allow answers in their own words.
- Often, asking people to tell a story about a “time when” a particular experience happened will evoke more detailed and candid responses than formal questions

# Tips on Interviews



- **General to specific.**
  - This order helps to avoid biasing later responses.
  - For example, if you first ask about what foods consumers like to buy at farmers' markets, they may already be thinking about those markets and answer subsequent questions in that context
- **Most to least important.**
  - Some respondents may have time constraints or become bored with the interview and end it early, and end early
- **Safest to riskiest.**
  - Best to open with a safe question to put the subject at ease and leave controversial or risky questions for later in the interview process

# Quantitative Methods



- Surveys are the most common method of gathering quantitative data.
  - They involve asking subjects a common set of questions, generally with short or close-ended answers
- As a rule of thumb, the order of questions in the survey instrument mirror those for interviews;
  - e.g., general to specific, most to least important and safest to riskiest. It is also customary to put demographic questions (age, race, gender, education) at the end of the survey.
- Having clear, straightforward instructions is critical in administering surveys

# Tips for Effective Surveys



- It is always good protocol to pilot test the survey with 5-10 volunteers to assure survey captures data you are looking for without creating excessive
- You will want to ask your volunteers to:
  - Take the survey;
  - Record how long it takes them to complete it;
  - Note any spelling or format errors;
  - Identify any questions that were difficult to understand or answer.
- Aim for surveys of less than 15 minutes
  - Only longer for those respondents very engaged in topic/issue

## Tools for Rapid Market Assessments



# The Case of Dot Surveys

- Dot Poster Surveys, also known as **Rapid Market Assessments**
  - Developed by Larry Lev and Garry Stephenson at Oregon State University to gather information from farmers' market patrons, and has many advantages.
- Simple to administer, responses are easily tallied, and possible to get a large set of responses in a short period of time.
- Respondents report that this method is faster to complete, more fun, and less intrusive than written surveys or face-to-face interviews.

# Dot Survey: Rapid Market Assessment



- Customers are given a strip of colored dots (one dot per question) to place on the corresponding answer. Examples of questions asked are:
  - How much did you spend at the farmers' market today?
  - Was the farmers' market your primary reason for coming downtown this morning?
  - Have you or do you plan to eat at a nearby restaurant or do additional shopping at a nearby business either before or after this market visit?
  - On average, if a specific item costs \$1.00 in the grocery store, how much would be willing to pay in the farmers' market for a similar product produced locally?

# Internet Surveys



- **Internet-based surveys have the advantages of:**
  - Relative ease of response--subjects respond when they wish
  - Cheaper to administer – no need to pay for travel to specific sites to conduct surveys
- **Internet surveys are increasingly popular**
  - Survey platforms may compile responses into a spreadsheet
  - It may be worth paying upfront for a site that supports robust backend features that facilitate data analysis.
- **Disadvantages include:**
  - Potential for a biased sample towards internet users, and
  - Difficulties limiting respondents



# Written Surveys



- In-person surveys generally sample the most convenient group, called **convenience sampling**
  - In-person surveys also run the risk of annoying people who came to shop, not take surveys.
- Mail surveys can be designed for a **representative sample**, but tend to have very low response rates.
  - **The Dillman Method**, consisting of an introductory letter, survey with addressed stamped return envelope, and reminder postcards is commonly used in mail surveys.
- Surveys of employees or key stakeholders should be well over 50%, while surveys of customers may be in the 20-40% range

# Phone Vs. Mail Surveys



Consider how these pros and cons translate to your target respondents and topics

- **Phone Surveys**

- Ability to sample selectively to reach sample quotas (a given percentage of females, for example);
- Quicker to complete and have available data;
- Survey caller can explain complex questions

- **Mail Survey**

- Lower cost;
- Ability to add visual graphics or longer questions

# More on Variables



- **Nominal.** These variables are categories that have no hierarchy (e.g., high/low, good/bad.).
  - Examples of these include the city/state where you live or were born, occupation, race or ethnicity, types of foods you eat
- **Ordinal.** Some degree of hierarchical order.
  - A very common question type that yields an ordinal variable is called a **Likert-scale**.
  - They are internally true (my ‘strongly agree’ is stronger than my ‘somewhat agree’), but your ‘strongly agree’ may be stronger than mine. Also, it is not possible to know the distance between each point of the scale

Strongly Disagree	Disagree	Neither Agree or Disagree	Agree	Strongly Agree
1	2	3	4	5

# Variable Choices



- **Interval and Ratio.** Distance between responses — interval and ratio — can be measured
  - Both internally (for a given person) and externally (between people).
  - In these cases, the ratio makes sense: someone who spent \$40 did spend twice as much as someone who spent \$20
- All variable types have pros and cons, so you just need to be thoughtful about what you want to learn
  - And, how you will analyze and present the data
- This is a case where drawing from others' survey instruments may be of value to your project

# Data Analysis: Qualitative Analysis Methods



- **Documentation** refers to the overall process of identifying recurring and important themes from your observations and interviews.
- What you are looking to track and identify are interactions between people, between people and ideas, and between people and places.
- You should also be looking for evidence of drivers:
  - Why people have made the choices they have
  - Keep in mind that people are different and there are many paths to particular outcomes and types of interaction that produce successful outcomes, so you should expect variation.

# Coding your Data

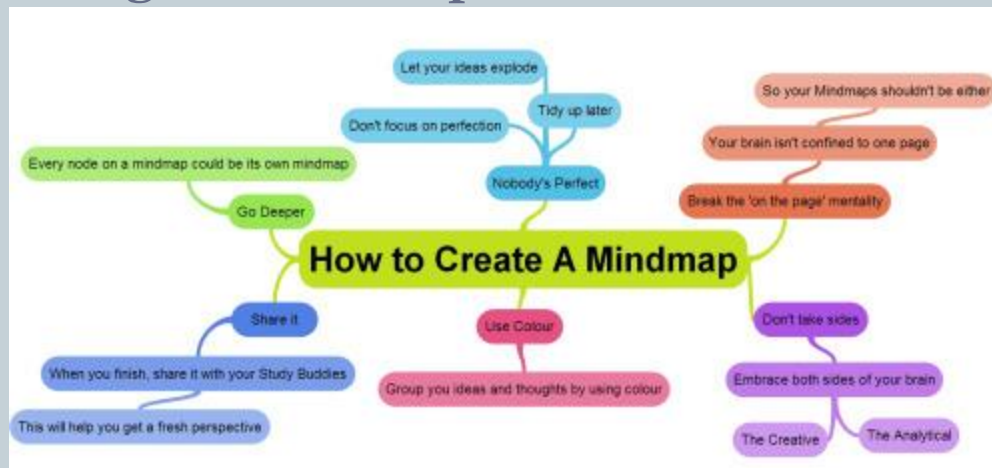


- The coding scheme you use should help you to group similar answers and enable you to draw conclusions
- Some of the codes will be pre-set based on interview questions, while others emerge from the analysis.
  - Suppose you asked farmers' market managers an open-ended question about why they had difficulty with vendor retention.
  - Each response will be different, but perhaps you can identify some overarching themes that you can then code.
  - For example, potential codes may include: market factors (e.g., not enough customer traffic); vendor factors e.g., (too much vendor competition); individual farm factors (e.g., farms too small to have adequate quantity of product and product mix).

# Drawing Relationships



- Many people use what is called a **mind map** to help them see these connections:
  - One begins to write on a blank piece of paper the recurring concepts and key themes that your research has identified.
  - Then, by drawing arrows to connect related themes, you often find strong relationships.



Source: London School of Economics



# Corroborating Findings



- Results should be triangulated or **corroborated** with other findings
  - If you find a trend in responses from farmers' market managers that does not appear in your farm responses, you need to consider alternative interpretations or explanations.
  - Do the farmers' market managers represent the same markets as farmers? Do market managers have access to different information? Are there representative samples of both groups?
- Essentially, if the different sources of information that you have do not corroborate each other, your team needs to do some deep digging and investigation into why this is the case

# A South Carolina Example

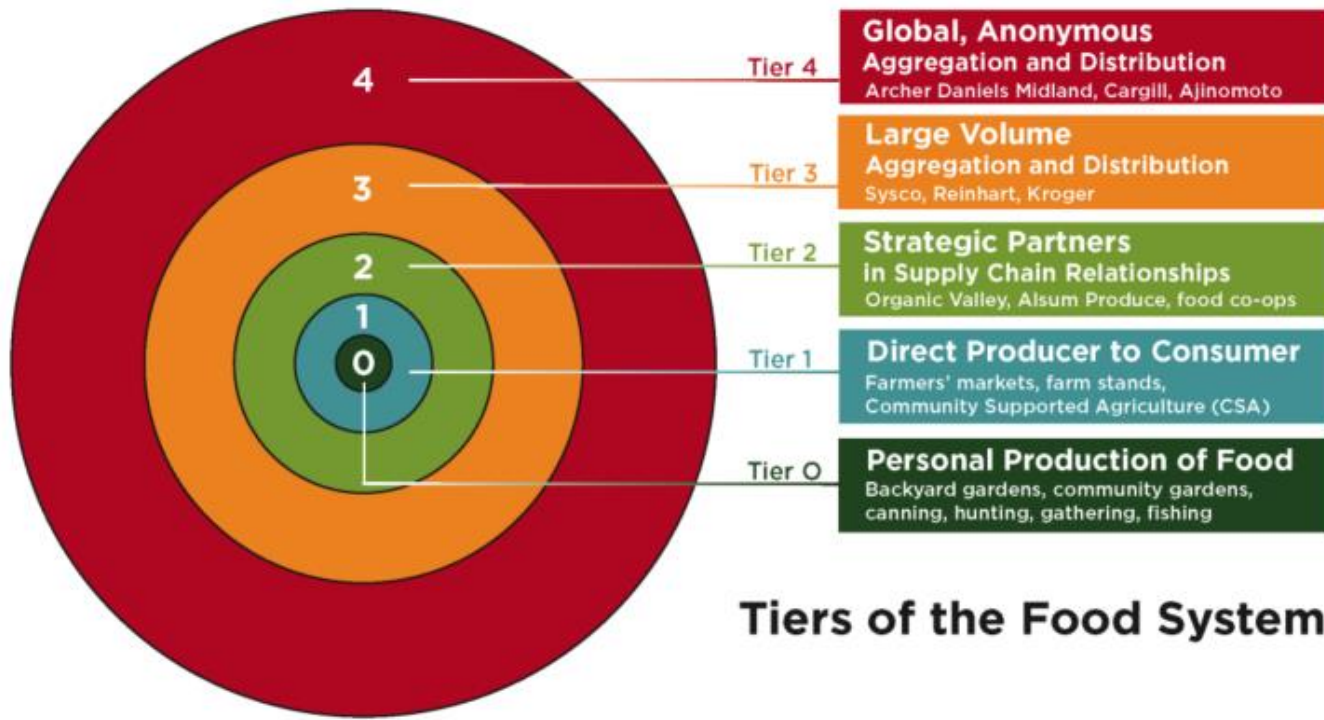


Figure 7: Diagram created at the University of Wisconsin Center for Integrated Agriculture Systems

South Carolina's 2013 *Making Small Farms into Big Business* provides a good example of using secondary data, before filling in information gaps with primary data collection.

For more information: [www.crcworks.org/scfood.pdf](http://www.crcworks.org/scfood.pdf)

# Details on South Carolina's Approach



- Compiled a variety of data from secondary sources:
  - USDA Economic Research Service, the Bureau of Labor Statistics, the Census of Agriculture, County Business Patterns, Centers for Disease Control and Prevention, National Center for Education Statistics, South Carolina Department of Natural Resources, and the National Hydrography dataset.
- Key informant interviews with over 150 practitioners, as well as a survey of specialty farmers
- By corroborating these secondary and primary data, the team came up with a strategic approach that addresses all levels of the state's food system (see Figure)

# Data Analysis: Quantitative Methods



- The first step in analyzing quantitative data is to calculate and report the **descriptive statistics**.
  - A spreadsheet program like Excel, considering variable type
- **Nominal variables**, it may be useful to calculate frequencies, or the volume of responses
  - For example, we collected data from 100 farmers, 46% of which identified fruit and vegetable production as their primary commodity.
- **Ordinal variables**, it is most common to report either the frequency or mean.
  - Mean may not show polarity or outliers so perhaps report both

# Quantitative Methods



- **Interval or ratio variables** you may choose to calculate the mean and median.
  - For example, if the variable is income and one household earns one million dollars per year and five households each earn \$20,000 per year, the **mean** household income in this sample of six households would be about \$183,000.
  - Reporting out the **range** (highest and lowest values) as well as the **median** may also be of interest.

If you have research findings that can be coded geographically, you may wish to bring your dataset to a Geographic Information System (GIS) specialist for mapping purposes

# Bivariate Analysis



- **Bivariate analyses** address questions about differences among various segments of population
  - Could be used to compare the means or frequencies of responses by men and women when asked “how much did you spend at this farmers market?”
- In general, you will want to compare means when conducting bivariate analysis of ratio/interval variables, compare frequencies when conducting bivariate analysis of ordinal or nominal variables.
- Be wary of making statistical inferences if a trained statistical researcher is not on your team

# Take Aways



- Primary data collection may be needed if no secondary data exist to answer your question(s).
  - Collection and analysis of data requires expertise and can be costly in time, money and effort.
- Qualitative methods provide a rich narrative of a subject's experiences and deeper understanding
- Quantitative methods often use survey results and provide an account of prevalence and correlation of important attributes on a larger sample.
- Mixed methods, using both qualitative and quantitative methods, complement each other well



# Take Aways



- Make sure methods will result in answers you need
- It is often useful to break a complex concept into its dimensions and then create questions and variables considering indicators important to that dimension
- Be thoughtful in choosing your sample
  - While representative samples may permit generalizing results to a larger population. . .
  - Convenience samples often provide useful information from knowledgeable stakeholders at a smaller cost in time and effort