6-2017

# Exploring Statistics Concepts with Student Driven Data and Polls

Matthew Patty
*Trinity University*, matthewapatty@gmail.com

**Understanding By Design**

**Unit Title:** Exploring Statistics Concepts with Student Driven Data & Polls

**Grade Level:** HS or College Introductory Statistics Course

**Author:** Matthew Patty

**Date of Publication**: June 2017

**Keywords:** Statistics, Graph, Polls, Survey, Probability, Valid, Prediction

**Abstract:** This unit, designed around the newly adopted Statisitics TEKS, will provide a strong framework for any introductory stats course, whether offered at the HS or collegiate level. The unit begins with a strong foundation in the study of statistics, and chellenges the students to design a series of surveys. The results from these surveys will actually drive the student through several of the next concepts in the course. Upon learning about sampling techniques, the students will actaully design a sampling method, carry out that method, analyze that sample data, and eventually make predictions on a population using thier sample data. This unit is designed to teach the basics of statistics through data that individually sparks interest for the students.

**Subject/Topic Area(s):** HS Mathematics

# Unit: Exploring Statistics Concepts with Student Driven Data & Polls
# Grade: HighSchool or College Introduction to Stats Course

## Stage 1: Desired Results

### Understandings

*Students will understand that...*

**Statistics are used to gather information, make decisions and predictions about populations.**

**Statistics allows us to make educated predictions about populations.**

**Statistics can be manipulated to promote specific opinions.**

**Graphical representations can often enhance communicating mathematical ideas and may provide support for predictions.**

**Outliers influence conclusions or predictions more than standard values.**

| Essential Questions | Knowledge & Skill |
|---|---|
| **When is a conclusion valid?**<br><br>**How accurate is a survey?**<br><br>**When is one type of graph the best?**<br><br>**Does it matter how you gather data?**<br><br>**Are there ethical reasons to accurately collect, analyze, and present data?**<br><br>**How does ethics play a role in the study of statistics?**<br><br>**Where is the center of my data?**<br><br>**How can one value misrepresent my data?**<br><br>**What does the spread of data communicate?**<br>**When is an outcome unlikely?** | *Although this unit could be used in any introduction to statistics course these Knoweldge & Skills statements are based on the Statistics TEKS.*<br><br>• **2.E - Formulate a meaningful question, determine the data needed to answer the question, gather the appropriate data, analyze the data, and draw reasonable conclusions.**<br><br>• **2.F - Communicate methods used, analyses conducted, and conclusions drawn for a data-analysis project through the use of one or more of the following: a written report, a visual display, an oral report, or a multi-media presentation.**<br><br>• **2.G - Critically analyze published findings for appropriateness of study design implemented, sampling methods used, or the statistics applied.**<br><br>• **2.C - Analyze generalizations made from observational studies, surveys, and experiments.** |

| | |
|---|---|
| **How does theoretical probability affect municipal growth?**<br><br>**How does probability affect our life?**<br><br>**When is a relationship strong enough to make predictions?**<br><br>**When can you make an accurate prediction?**<br><br>**Is everything in life linear?** | • **5.A - Determine probabilities, including the use of a two-way table.**<br><br>• **7.A - Analyze scatterplots for patterns of linearirty.** |

# Stage 2: Assessment Evidence

Performance Task

**Students will create a question that can be addressed with a series of polls; one bi-variate qualitative poll, one bi-variate quantitative poll, and a final poll with a quantitative question as well as a qualitative question . They will then design a sampling method, carry out that sampling method, and then use the resulting data to drive all instruction for the following units. The performance task will be the collection of data at the conclusion of the units described below.**

"Statistics show up all around us, and it only makes sense that we use the mathematics in our real world as we study this course. You will begin the year by designing a series of questions that you will work to make predictions about this semester. After deciding on the questions you will design a sampling technique that will provide you with the most accurate responses and then you will carry out the data collection. After collecting the data we will begin analyzing this data as we explore the rest of the topic in the course."

Other evidence:

*(quizzes, tests, academic prompts, self-assessments, etc.*
*note – these are usually included where appropriate in Stage 3 as well)*

**Throughout the data collection and analysis stages, the students must present their findings to their peers. These will follow structured feedback forums as well as informal discussions.**

**There will be daily exit slips collected via Google Forms.**

**Plickers will also be used on a daily basis to ensure students are following along during the direct teach sessions.**

# Stage 3: Learning Activities

*(Steps taken to get students to answer Stage 1 questions and complete performance task)*

## Phase 1: Designing a Good Poll (1.5 hours)

**[EQ - When is a conclusion valid?]**

**[EQ - Are there ethical reasons to accurately collect, analyze, and present data?]**

**Introduction**

- Using a concept attainment activity students will determine the definition of the following terms;

| | | | |
|---|---|---|---|
| Statistic | Parameter | | |
| Sample | Population | Quantitative Data | Qualitative Data |
| Poll | Census | Continuous data | Discrete Data |

**Potential Pitfalls in Anlyzing Data**

- Students will complete a card sort activity in which they match the pitfall title with the description.

**Creating Your Questions**

- Students must decide and design their questions. Once completed they can submit their work here.

## Phase 2: How to Collect Data (1.5 hours)

**[EQ - Does it matter how you gather data?]**

**Introduction**

- Showcase several biased polls and have students determine where the bias comes from. These should be examples such as;
    - Online Polls for access to technology
    - Online shopping taken at a mall
    - Chocolate is good for health sponsored by Mars Inc.

**Sampling Methods**

- Complete the Does Beyonce Write Her Own Songs actvity in which students will determine the differences between sampling methods, the benefits of a large sample, and the need for a valid sampling technique.

**Designing Your Collection Technique:**

- Students must brainstorm and determine how they will collect their data. Google Forms will be encouraged and a very brief overview of that tool will be provided outside of class time.

## Phase 3: Summarizing & Graphing Data (3 hours)

**[EQ - What type of graph is best?]**

**Frequency Distributions (Quantitative & Qualitative):**

- Students will be introduced to the concepts of a frequency distribution table using an example of the last 2 digits of their cell phone number.
- The students will then be asked to create their own frequency tables of their data.

**Histograms (Quantitative)**

- The phone number example will be continued to create a histogram of the data.
- The students will then be asked to create their own histogram from one of their frequency tables. This must be the quantitative data, however the students must determine this on their own.

**Graphical Representations (All)**

- Students will be taken through a series of graphs representing the same data set it will be clear that some do a much better job of comunicating the desired result.

## Phase 4: Describing, Exploring, & Comparing Data (4.5 hrs)

**[EQ - Where is the center of my data?]**
**[EQ - How can one value represent my data?]**
**[EQ - How spread out is my data?]**
**[EQ - How abnormal are you?]**

**Introduction**

- Give students the example of going to a supermarket and seeing three registers to choose from (A,B,C). each line has three people in line, each with a basket of itmes, however, each basket has a different number of items. The diagram looks like below;

| Register | # of Items in Basket 1 | # of Items in Basket 2 | # of Items in Basket 3 |
|---|---|---|---|
| A | 6 | 6 | 6 |
| B | 4 | 7 | 7 |
| C | 1 | 3 | 14 |

  Ask the students which register they would get into line with. Collect and display their responses via Plickers.
- Discuss with the students how beneficial it would be if you could do a quick calculation and make an educated choice on which line you should choose. (This is statistics!!!) This should lead to a natural transition into the next conversation.

**Measures of Center (Quantitative/Qualitative)**

- Using the supermarket example extend the table and calculate the mean, median, and mode.

    - Include greek notation to ensure success in the later portion of the course.
- Have students calculate Mean, Median, Mode, MidRange of their samples (Quantitative) (Quan/Qual)
- Have students complete a written comparison about their Measures of Centers and justify which one best represents their data.

**Measures of Spread (Quantitative/Qualitative)**

- Continue supermarket Conversation to discuss wanting to make an even more informed decision about which line to step into. Discuss that the option is no longer a simple supermarket line, but a financial investment.
    - Include conversation about units and include how and why standard deviation is a better measure for practical use, but variance is needed for the formal theory of statistics due to the square root function not being bijective.
- Have students calculate Range, Standard Deviation, and Variance of their samples (Quantitative) (Quan/Qual)
- Have students complete a written comparison about their Measures of Spread and justify which one best represents their data.

**Measures of Relative Standing (Z-Scores)**

- Discuss the unfairness and disappointment of ordering a pepperoni pizza, but getting a slice that only has 2 pepperonis. Transfer this atrocity to the feeling of sitting down to enjoy a chocolate chip cookie only to discover that it only has 11 choclate chips! How unfair is that?!
- Using this example as a spring board discuss that we can actually quantify how unfair (or how unlikely) situations like this are by using the calculation of a z-score.
    - Include conversation about how the z-score is measuring the number of standard deviations the single data point is from the average. Add that this definition has tremendous more effect when data is normally distributed.
        - If time allows touch on Emprical Rule.
- Have students calculate z-scores for 5 random points from each of their data sets,
- Have students complete a written summary reflecting on at least one point (and it's corresponding z-score) from each data set.

**BoxPlots (Quantitative/Qualitative)**

- Using the chocolate chip data, create three box plots to compare the different data sets.
    - Discuss the characteristics of a box plot, the 5 number summary, and include a brief conversation on outliers.
- Have students calculate 5 number summaries, and box plots of their samples (Quan/Qual)
- Have students complete a written comparison about their three box plots comparing the data and making a valid conclusion about them.

## Phase 5: Probability (4.5 hrs)

**[EQ - Why does Las Vegas exist?]**
**[EQ - What are the chances?]**

**Introduction**

- With three short examples including roulette, cards, and dice, cover the basic probability

vocabulary including event, simple event, sample space, compliment and disjoint events.

- Briefly explain the three types of probability (*Relative Frequency Approximation of Probability (Observed), Classical Approach (Theoretical) to Probability, Subjective Probability)* and then complete a short 4 corners activity where the students must read an example and move to identify which type of probability each situation is.
    - At the conclusion of the discussion ensure the students have an understanding of the Law of Large Numbers.

**Discrete Probability Distribution**

- Create the discrete probability table for the gambling situations described in the introduction.
- Next, create a 2-way table with fake data (to mock the student's data) and build a discrete probability table for this data.
- Have students create a discrete probability table with their own 2-way table created from qualitative data.

**Addition/Subtraction Rule (Qualitative)**

- Using the created 2-way table with fake data, discuss concepts of Disjoint and non disjoint events. Proceed through this discussion with the "I do, We do, You do" pattern.
    - Include formal language and notation during conclusion.

**Multiplication Rule (Qualitative)**

- Using the same 2-way table continue with the examples to teach independent events, dependent events, basic compliments, and advanced compliments ("At least One.")Proceed through this discussion with the "I do, We do, You do" pattern.
    - Include formal language and notation during conclusion.
- Have students create at least 2 questions from their data that can be solved with these concepts. When submitting these questions they must include worked out solutions.
    - 1 question must include disjoint probabilities.
    - 1 question must include non-disjoint probabilities.
    - 1 question must include dependent conditional probabilities.
    - 1 question must include independent condtional probabilities.
    - 1 question must include basic compliments.
    - 1 question must include advanced compliments.

## Stage 6: Correlation & Regression (3 hrs)

**[EQ - When is a relationship strong enough to make predictions?]**
**[EQ - When can you make an accurate prediction?]**
**[EQ - Is everything in life linear?]**

**Introduction:**

- To introduce concepts of general correlation (and to implicitly talk about how correlation does not imply causation) have students explore Spurious Correlations website. Have them pick one graph and complete a brief reflection identifying what the graph is attempting to

imply, and whether or not they agree with that implied conculsion.

- Next to transition into linear correlation use a series of basic real world examples to introduce Positive, Negative, No, and Non-Linear Correlations.

**Correlation Linear Coefficient**

- Introduce this new variable with a take on Concept Attainment. Returning to the real world examples used to introduce linear correlation, but this time add the correlation coefficient value on the screen. Have the students determine the parameters, and the definition of the correlation coefficient.

**Regression (Quantitative/Quantitative)**

- Using a Fake Data example, introduce the basics of how to calculate a Regression Line.

**Least Squares Method**

- Studnets will explore Geogebra/Desmos applet visualizing the least squares technique.

**Making Predictions based on the Regression Line**

- Using our fake data, we will calculate, and plot the regression line. Then use this data to make predictions about life.
- Have students must create a scatter plot with their data. Have them include the calculated and drawn regression line. They will then use this line to create a prediction.

**Life is Not Linear**

- Based on the chapter in the Jordan Ellenburg text, How Not to Be Wrong: The Power of Mathematical Thinking, entitled “Straight Locally, Curved Globally” discuss when and where linear thinking may faulter.

## Calendar and Assignment - Math 1442 with UdB Integration

**This is a tentative schedule and is subject to change**

| Date | Class | Topic |
|---|---|---|
| | C1 | Phase 1: Designing a Good Poll |
| | C2 | Phase 2: How to Collect Data |
| | C3 | Phase 3: Summarizing & Graphing Data |
| | C4 | |
| | C5 | Phase 4: Describing, Exploring, & Comparing Data |
| | C6 | |
| | C7 | Phase 5: Probability |
| | C8 | |
| | C9 | |
| | C10 | Phase 6: Correlation & Regression |
| | C11 | |
| | C12 | Presentations |
| | C13 | Buffer |
| | C14 | Counting & Binomial Probability Distributions |
| | C15 | Review of Chapters 1 - 5 |
| | C16 | **Test # 2 Over Chapters 1 - 5** |
| | C17 | The Standard Normal Distribution |
| | C18 | Applications of Normal Distributions |
| | C19 | The Central Limit Theorem |
| | C20 | Estimating a Population Proportion |
| | C21 | Review of Chapters 6 - 7 |
| | C22 | **Test # 3 Over Chapters 6 - 7** |
| | C23 | Basics of Hypothesis Testing |
| | C24 | Testing a Claim about a Proportion |
| | C25 | Testing a Claim about a Mean |
| | C26 | Two Proportions |
| | C27 | Two Means: Independent Samples |
| | C28 | Correlation & Regression Revisited (P-Values) |
| | C29 | |
| | C30 | Review of Chapters 8 - 10 |
| | | **Final Exam Over Chapters 8 - 10**** |

**Your Data Driving Your Statistics Course**

Statistics show up all around us, and it only makes sense that we use the mathematics in our real world as we study this course. You will begin the year by designing a series of questions that you will work to make predictions about this semester. After deciding on the questions you will design a sampling technique that will provide you with the most accurate responses and then you will carry out the data collection. After collecting the data we will begin analyzing this data as we explore the rest of the topic in the course.

**Stage 1: Desgining A Good Poll**

Your task begins with designing 3 short polls. Each poll must have at least two questions.

- One poll must contain at least two qualitative questions.
    - *What gender do you associate yourself with?*
    - *What political party do you associate yourself with?*
- One poll must contain at least two quantitative question.
    - *In years, how old are you?*
    - *How many pets do you currently have?*
- One poll must contain at least one qualitative and one quantitative question.
    - *What state were you born in?*
    - *How many US states have you visited?*

When you are ready to submit **Stage 1** please complete this Google Form.

**Stage 2: How to Collect Data**

Now that you have your polls, let's collect data!

- You must collect at least 100 responses for each of your polls. The same person may respond to all three polls questions if they would like. You do not have to complete these polls on paper and pencil, if you would like to use Google Forms you may - here is an example of a digital poll.
    - There are some advantages to collecting these responses digitally, however your personal devices may have to be used to do this.
- You must also use an appropriate sampling technique to collect these samples.

When you are ready to submit **Stage 2** please complete this Google Form.

**Stage 3: Summarizing & Graphing Data**

We will begin analyzing your individual data with some Distribution Frequency Tables

- You may create these electronically or on paper.
    - There are some advantages to collecting these responses digitally, however your personal devices may have to be used to do this.
- You must create at least 2 frequency tables
    - One must come from your qualitative data
    - One must come from your quantitative data
        - You may create your own class limits, boundaries, and midpoints.
        - You may ceate your own number of classes, but I would reccomend 4 to 6.
        - Please use the standard calculation for class width;
            - Class Width = ((Max data value - Min data value) / number of classes)
                - Remember round **any** decimal place up.
    - One of the frequency tables must include both a relative frequency distribution, and a cummulative frequency distribution.
- You must create at least 1 histogram
    - This must come from your _____ data (You have to determine which data set you can use)
    - Use the same class limits, boundaries, and midpoints as you did for the frequency tables.

When you are ready to submit **Stage 3** please add your work below (this can be images, computer generated, or hand drawn)

**Frequency Tables:**

**Histogram:**

**Stage 4: Describing, Exploring, & Comparing Data**

- As discussed in class, you must calculate the mean, median, mode, midrange, range, variance, and standard deviation of your data.
    - You should be doing this on your qualitative/quantitative data, so your measurements will be split into categories. This will allow for comparisions to happen naturally.
- Please include your calculated values below. There is no need to include the entire data or show your calculations;

Measures of center:

Measures of Spread:

**Stage 4: Describing, Exploring, & Comparing Data**

- As discussed in class, you must calculate z-scores for up to five random points from each category of your data.
- Please display your calculated values below. Include the original data value, the mean the corresponding z-scores

Z-Scores:

**Stage 4: Describing, Exploring, & Comparing Data**

- As discussed in class, you must create and compare three box plots based on your qualitative/quantitative data.
    - Use your qualitative categories to split your data into three categories if this is not naturally done already.
- Please include a sketch of your three box plots below.
    - If you have created these electronically you may include screen shots below.

Box Plots

When you are ready to submit **Stage 4** please complete this Google Form.

**Stage 5: Probability**

- As discussed in class, you must create a two way table including the totals from your qualtitative questions.
- Please include your 2-way table below;
    - If you have created these electronically you may include screen shots below.

2-Way Table

**Stage 5: Probability**

- As discussed in class, you must create a discrete probability distribution from your 2-way table
- Please include your discrete probability distribution below;
    - If you have created these electronically you may include screen shots below.

Discrete Probability Distribution

**Stage 5: Probability**

- As discussed in class, you must create at least 6 questions from your data that can be solved using the theory from class.
    - Include both the question, and the worked out solution.

Disjoint Probability

Non-disjoint Probability

**Stage 5: Probability**

- As discussed in class, you must create at least 6 questions from your data that can be solved using the theory from class.
    - Include both the question, and the worked out solution.

Dependent Probability

Independent Conditional Probability

**Stage 5: Probability**

- As discussed in class, you must create at least 6 questions from your data that can be solved using the theory from class.
    - Include both the question, and the worked out solution.

Basic Compliment

Advanced Compliment

**Stage 6: Correlation & Regression**

- As discussed in class, you must create a scatter plot of your bivartiate quantitative data. We will determine how strong of a linear realtionship exists within your data.
- Please include your scatter plot below
    - Include the correlation coefficient, and the regression line.
        - You may determine the equation of this line electronically, but please include the equation in your scatter plot.
    - If you have created this electronically you may include screen shots below.

Scatter Plot

When you are ready to submit **Stage 5** please complete this Google Form.

# Does Beyonce write her own lyrics?



*"Crazy in Love' was really hard to write because there was so much going on … I mean, I had written — what? —seven, eight number one songs with Destiny's Child, in a row." Beyonce, Vanity Fair*

We can use statistics to help determine whether or not Beyonce wrote the song "Crazy in Love". If we can find the average word length from the song, we can compare it to the average word length for songs that we know for sure were written by Beyonce.

1. ___Convenience Sampling_____
    a. Quickly circle a random sample of 5 words from the song.
    b. How many letters in each word?
    c. What is the average word length of your sample? _____.
    d. Put your average on the dotplot on the white board at the front of the room.
    e. Copy the class dotplot below.

2. _____Systematic Sampling_____
    a. Using the random number generator, generate a number between 20 and 40. Then select every "nth" word. (Example if you picked 25 you would pick number 25, 50, 75, excetra…).
    b. How many letters in each word?
    c. What is the average word length of your sample? _____ .
    d. Put your average on the dotplot on the white board at the front of the room.
    e. Copy the class dotplot below.

3. _____Stratification Sampling_____
    a. First we will define our stratas as the leading letter of each word. Then we will orgainze all of our data into their seperate strata. Notice not all strata are the same size, but no element can belong to more than one strata.
    b. Now we will pick one word per strata. To choose the one words per strata we will use a random number generator to determine which words we will select.
    c. How many letters in each word?
    d. What is the average word length of your sample?
    e. Put your average on the dotplot on the white board at the front of the room.
    f. Copy the class dotplot below.

4. _____Cluster Sampling_____
    a. Using the lyric list we will compile clusters of the lyrics grouping every 10th word. Notice not all clusters are the same size, but no element can belong to more than one cluster.
    b. Now we will randomly select one of the 25 clusters by using a random number generater. We will then choose all members of the randomly selected cluster for our calculation.
    c. How many letters in each word?

d. What is the average word length of your sample?
e. Put your average on the dotplot on the white board at the front of the room.
f. Copy the class dotplot below.

5. On a first glance stratified sampling and cluster sampling seem similar, but there are a few small but very important differences, what are they?

6. _____Simple Random Sample_____
   a. Using a random number generator, create a sample of 5 words.
   b. What is the average word length of your sample?
   c. Put your average on the dotplot on the white board at the front of the room.
   d. Copy the class dotplot below.

7. _____Simple Random Sample_____
   a. Using a random number generator, create a sample of 10 words.
   b. What is the average word length of your sample?
   c. Put your average on the dotplot on the white board at the front of the room.
   d. Copy the class dotplot below

8. What happens to the dotplot when we increase the sample size?

9. It is a well known fact that Beyonce wrote the lyrics for all of the Destiny's child songs. The average word length for these songs is 3.64 letters. Based on your samples, do you have good evidence that Beyonce did not write the lyrics for "Crazy in Love". Explain.

# Crazy in Love

I look and stare so deep in your eyes
I touch on you more and more every time
When you leave I'm begging you not to go
Call your name two or three times in a row
Such a funny thing for me to try to explain
How I'm feeling and my pride is the one to blame
'Cuz I know I don't understand
Just how your love can do what no one else can

Got me looking so crazy right now, your love's
Got me looking so crazy right now (in love)
Got me looking so crazy right now, your touch
Got me looking so crazy right now (your touch)
Got me hoping you'll page me right now, your kiss
Got me hoping you'll save me right now
Looking so crazy in love's
Got me looking, got me looking so crazy in love

When I talk to my friends so quietly
Who he think he is? Look at what you did to me
Tennis shoes, don't even need to buy a new dress
If you ain't there ain't nobody else to impress
The way that you know what I thought I knew
It's the beat my heart skips when I'm with you
But I still don't understand
Just how the love your doing no one else can

I'm Looking so crazy in love's
Got me looking, got me looking so crazy in love

Got me looking, so crazy, my baby
I'm not myself, lately I'm foolish, I don't do this
I've been playing myself, baby I don't care
'Cuz your love's got the best of me
And baby you're making a fool of me
You got me sprung and I don't care who sees
'Cuz baby you got me, you got me, so crazy baby
HEY!

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | I | 51 | and | 101 | your | 151 | so | 201 | I'm | 251 | this |
| 2 | look | 52 | my | 102 | touch | 152 | quietly | 202 | with | 252 | I've |
| 3 | and | 53 | pride | 103 | Got | 153 | Who | 203 | you | 253 | been |
| 4 | stare | 54 | is | 104 | me | 154 | he | 204 | But | 254 | playing |
| 5 | so | 55 | the | 105 | looking | 155 | think | 205 | I | 255 | myself |
| 6 | deep | 56 | one | 106 | so | 156 | he | 206 | still | 256 | baby |
| 7 | in | 57 | to | 107 | crazy | 157 | is | 207 | don't | 257 | I |
| 8 | your | 58 | blame | 108 | right | 158 | Look | 208 | understand | 258 | don't |
| 9 | eyes | 59 | 'Cuz | 109 | now | 159 | at | 209 | Just | 259 | care |
| 10 | I | 60 | I | 110 | (your | 160 | what | 210 | how | 260 | 'Cuz |
| 11 | touch | 61 | know | 111 | touch) | 161 | you | 211 | the | 261 | your |
| 12 | on | 62 | I | 112 | Got | 162 | did | 212 | love | 262 | love's |
| 13 | you | 63 | don't | 113 | me | 163 | to | 213 | your | 263 | got |
| 14 | more | 64 | | 114 | hoping | 164 | me | 214 | doing | 264 | the |
| 15 | and | understand | | 115 | you'll | 165 | Tennis | 215 | no | 265 | best |
| 16 | more | 65 | Just | 116 | page | 166 | shoes | 216 | one | 266 | of |
| 17 | every | 66 | how | 117 | me | 167 | don't | 217 | else | 267 | me |
| 18 | time | 67 | your | 118 | right | 168 | even | 218 | can | 268 | And |
| 19 | When | 68 | love | 119 | now, | 169 | need | 219 | I'm | 269 | baby |
| 20 | you | 69 | can | 120 | your | 170 | to | 220 | Looking | 270 | you're |
| 21 | leave | 70 | do | 121 | kiss | 171 | buy | 221 | so | 271 | making |
| 22 | I'm | 71 | what | 122 | Got | 172 | a | 222 | crazy | 272 | a |
| 23 | begging | 72 | no | 123 | me | 173 | new | 223 | in | 273 | fool |
| 24 | you | 73 | one | 124 | hoping | 174 | dress | 224 | love's | 274 | of |
| 25 | not | 74 | else | 125 | you'll | 175 | If | 225 | Got | 275 | me |
| 26 | to | 75 | can | 126 | save | 176 | you | 226 | me | 276 | You |
| 27 | go | 76 | Got | 127 | me | 177 | ain't | 227 | looking, | 277 | got |
| 28 | Call | 77 | me | 128 | right | 178 | there | 228 | got | 278 | me |
| 29 | your | 78 | looking | 129 | now | 179 | ain't | 229 | me | 279 | sprung |
| 30 | name | 79 | so | 130 | Looking | 180 | nobody | 230 | looking | 280 | and |
| 31 | two | 80 | crazy | 131 | so | 181 | else | 231 | so | 281 | I |
| 32 | or | 81 | right | 132 | crazy | 182 | to | 232 | crazy | 282 | don't |
| 33 | three | 82 | now, | 133 | in | 183 | impress | 233 | in | 283 | care |
| 34 | times | 83 | your | 134 | love's | 184 | The | 234 | love | 284 | who |
| 35 | in | 84 | love's | 135 | Got | 185 | way | 235 | Got | 285 | sees |
| 36 | a | 85 | Got | 136 | me | 186 | that | 236 | me | 286 | 'Cuz |
| 37 | row | 86 | me | 137 | looking, | 187 | you | 237 | looking, | 287 | baby |
| 38 | Such | 87 | looking | 138 | got | 188 | know | 238 | so | 288 | you |
| 39 | a | 88 | so | 139 | me | 189 | what | 239 | crazy, | 289 | got |
| 40 | funny | 89 | crazy | 140 | looking | 190 | I | 240 | my | 290 | me, |
| 41 | thing | 90 | right | 141 | so | 191 | thought | 241 | baby | 291 | you |
| 42 | for | 91 | now | 142 | crazy | 192 | I | 242 | I'm | 292 | got |
| 43 | me | 92 | (in | 143 | in | 193 | knew | 243 | not | 293 | me, |
| 44 | to | 93 | love) | 144 | love | 194 | It's | 244 | myself, | 294 | so |
| 45 | try | 94 | Got | 145 | When | 195 | the | 245 | lately | 295 | crazy |
| 46 | to | 95 | me | 146 | I | 196 | beat | 246 | I'm | 296 | baby |
| 47 | explain | 96 | looking | 147 | talk | 197 | my | 247 | foolish, | 297 | HEY |
| 48 | How | 97 | so | 148 | to | 198 | heart | 248 | I | | |
| 49 | I'm | 98 | crazy | 149 | my | 199 | skips | 249 | don't | | |
| 50 | feeling | 99 | right | 150 | friends | 200 | when | 250 | do | Credit: Gabe Yonker | |
| | | 100 | now, | | | | | | | | |

## Stratified Sampling

| | | |
|---|---|---|
| A | a a a a ain't ain't and and And and and at | 12 |
| B | baby baby baby baby baby beat been begging best blame But buy | 12 |
| C | Call can can can care care crazy crazy crazy crazy crazy crazy crazy crazy crazy crazy Cuz Cuz Cuz | 19 |
| D | deep did do do doing don't don't don't don't don't don't dress | 12 |
| E | else else else even every explain eyes | 7 |
| F | feeling fool foolish for friends funny | 6 |
| G | go Got got Got Got got Got got got Got got Got Got Got got Got | 16 |
| H | he he heart HEY hoping hoping how how How | 9 |
| I | I I I I I I I I I I I I I'm I'm I'm I'm I'm I'm I've If impress in in in in in in in is is It's | 30 |
| J | Just Just | 1 |
| K | kiss knew know know | 4 |
| L | lately leave Look look looking Looking Looking looking looking looking looking looking looking looking love love love love love love's love's love's love's | 24 |
| M | making me me me me me me me me me me me me me me me me me me me more more my my my my myself myself | 29 |
| N | name need new no no nobody not not now now now now now now | 14 |
| O | of of on one one one or | 7 |
| P | page playing pride | 3 |
| Q | quietly | 1 |
| R | right right right right right right row | 7 |
| S | Such save sees shoes skips so so so so so so so so so so so so sprung stare still | 20 |
| T | talk Tennis that The the the the the there thing think this thought three times time | 16 |
| U | | 0 |
| V | | 0 |
| W | When to to to to to to to to touch touch touch try two understand understand way what what what When when Who who with | 25 |
| X | | 0 |
| Y | you you you you You you you you you you you'll you'll you're your your your your your your your your your | 22 |
| Z | | 0 |

## Cluster Sampling

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 1 | I | 51 | and | 101 | your | 151 | so | 201 | I'm | 251 | this |
| 2 | look | 52 | my | 102 | touch | 152 | quietly | 202 | with | 252 | I've |
| 3 | and | 53 | pride | 103 | Got | 153 | Who | 203 | you | 253 | been |
| 4 | stare | 54 | is | 104 | me | 154 | he | 204 | But | 254 | playing |
| 5 | so | 55 | the | 105 | looking | 155 | think | 205 | I | 255 | myself |
| 6 | deep | 56 | one | 106 | so | 156 | he | 206 | still | 256 | baby |
| 7 | in | 57 | to | 107 | crazy | 157 | is | 207 | don't | 257 | I |
| 8 | your | 58 | blame | 108 | right | 158 | Look | 208 | understand | 258 | don't |
| 9 | eyes | 59 | 'Cuz | 109 | now | 159 | at | 209 | Just | 259 | care |
| 10 | I | 60 | I | 110 | (your | 160 | what | 210 | how | 260 | 'Cuz |
| 11 | touch | 61 | know | 111 | touch) | 161 | you | 211 | the | 261 | your |
| 12 | on | 62 | I | 112 | Got | 162 | did | 212 | love | 262 | love's |
| 13 | you | 63 | don't | 113 | me | 163 | to | 213 | your | 263 | got |
| 14 | more | 64 | understand | 114 | hoping | 164 | me | 214 | doing | 264 | the |
| 15 | and | 65 | Just | 115 | you'll | 165 | Tennis | 215 | no | 265 | best |
| 16 | more | 66 | how | 116 | page | 166 | shoes | 216 | one | 266 | of |
| 17 | every | 67 | your | 117 | me | 167 | don't | 217 | else | 267 | me |
| 18 | time | 68 | love | 118 | right | 168 | even | 218 | can | 268 | And |
| 19 | When | 69 | can | 119 | now, | 169 | need | 219 | I'm | 269 | baby |
| 20 | you | 70 | do | 120 | your | 170 | to | 220 | Looking | 270 | you're |
| 21 | leave | 71 | what | 121 | kiss | 171 | buy | 221 | so | 271 | making |
| 22 | I'm | 72 | no | 122 | Got | 172 | a | 222 | crazy | 272 | a |
| 23 | begging | 73 | one | 123 | me | 173 | new | 223 | in | 273 | fool |
| 24 | you | 74 | else | 124 | hoping | 174 | dress | 224 | love's | 274 | of |
| 25 | not | 75 | can | 125 | you'll | 175 | If | 225 | Got | 275 | me |
| 26 | to | 76 | Got | 126 | save | 176 | you | 226 | me | 276 | You |
| 27 | go | 77 | me | 127 | me | 177 | ain't | 227 | looking, | 277 | got |
| 28 | Call | 78 | looking | 128 | right | 178 | there | 228 | got | 278 | me |
| 29 | your | 79 | so | 129 | now | 179 | ain't | 229 | me | 279 | sprung |
| 30 | name | 80 | crazy | 130 | Looking | 180 | nobody | 230 | looking | 280 | and |
| 31 | two | 81 | right | 131 | so | 181 | else | 231 | so | 281 | I |
| 32 | or | 82 | now, | 132 | crazy | 182 | to | 232 | crazy | 282 | don't |
| 33 | three | 83 | your | 133 | in | 183 | impress | 233 | in | 283 | care |
| 34 | times | 84 | love's | 134 | love's | 184 | The | 234 | love | 284 | who |
| 35 | in | 85 | Got | 135 | Got | 185 | way | 235 | Got | 285 | sees |
| 36 | a | 86 | me | 136 | me | 186 | that | 236 | me | 286 | 'Cuz |
| 37 | row | 87 | looking | 137 | looking, | 187 | you | 237 | looking, | 287 | baby |
| 38 | Such | 88 | so | 138 | got | 188 | know | 238 | so | 288 | you |
| 39 | a | 89 | crazy | 139 | me | 189 | what | 239 | crazy, | 289 | got |
| 40 | funny | 90 | right | 140 | looking | 190 | I | 240 | my | 290 | me, |
| 41 | thing | 91 | now | 141 | so | 191 | thought | 241 | baby | 291 | you |
| 42 | for | 92 | (in | 142 | crazy | 192 | I | 242 | I'm | 292 | got |
| 43 | me | 93 | love) | 143 | in | 193 | knew | 243 | not | 293 | me, |
| 44 | to | 94 | Got | 144 | love | 194 | It's | 244 | myself, | 294 | so |
| 45 | try | 95 | me | 145 | When | 195 | the | 245 | lately | 295 | crazy |
| 46 | to | 96 | looking | 146 | I | 196 | beat | 246 | I'm | 296 | baby |
| 47 | explain | 97 | so | 147 | talk | 197 | my | 247 | foolish, | 297 | HEY |
| 48 | How | 98 | crazy | 148 | to | 198 | heart | 248 | I | | |
| 49 | I'm | 99 | right | 149 | my | 199 | skips | 249 | don't | | |
| 50 | feeling | 100 | now, | 150 | friends | 200 | when | 250 | do | | |