THE UNIVERSITY OF TEXAS AT DALLAS

A/B Testing and Regression Analysis

Vatsal Maru, PhD
JSOM 4.414

SEO Measurement

- Current performance.
- Content coverage
- Keyword performance

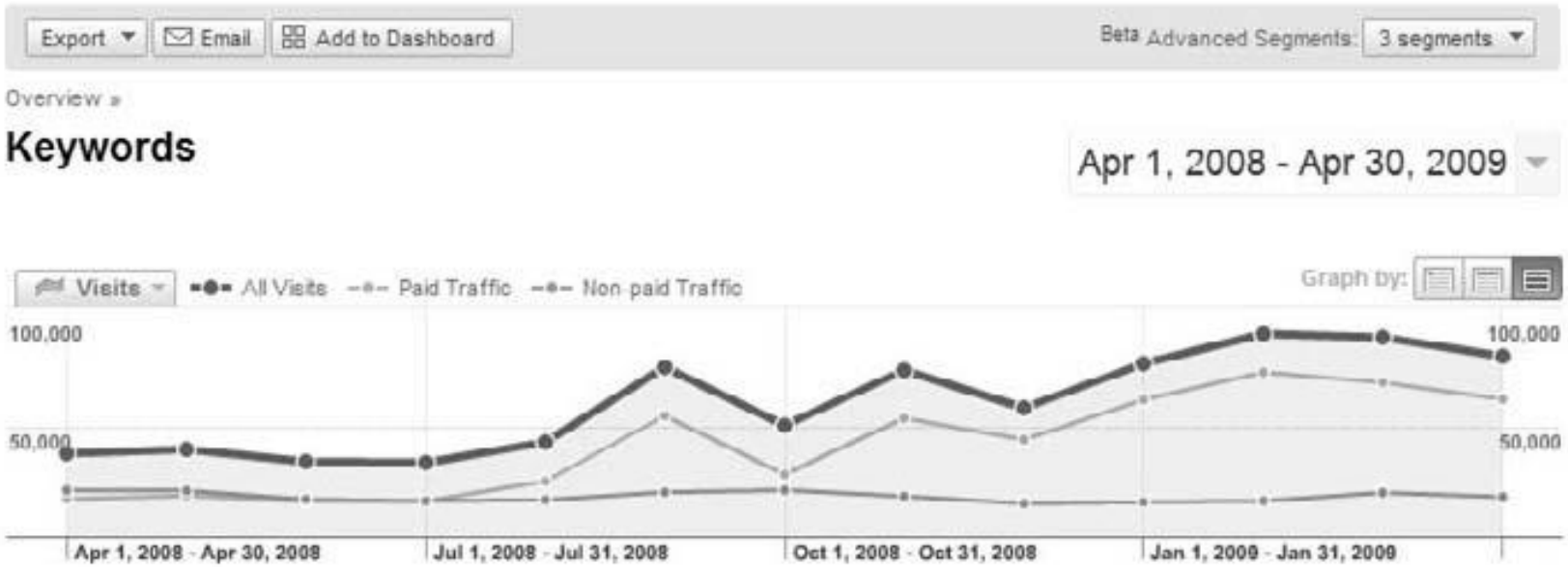
Kaushik, A. (2009). *Web analytics 2.0: The art of online accountability and science of customer centricity*. John Wiley & Sons.

Performance

- Can see key statistics as well as understand the performance of the organic search traffic by looking at metrics such as Bounce Rate and CTR
- You can also click the Goal Conversion tab and quickly assess the value of organic traffic

Kaushik, A. (2009). *Web analytics 2.0: The art of online accountability and science of customer centricity*. John Wiley & Sons.

Kaushik, A. (2009). *Web analytics 2.0: The art of online accountability and science of customer centricity*. John Wiley & Sons.



Kaushik, A. (2009). *Web analytics 2.0: The art of online accountability and science of customer centricity*. John Wiley & Sons.

Coverage

- Indexing by search engine
- Your explicit goal is to get your website indexed properly by the search engine

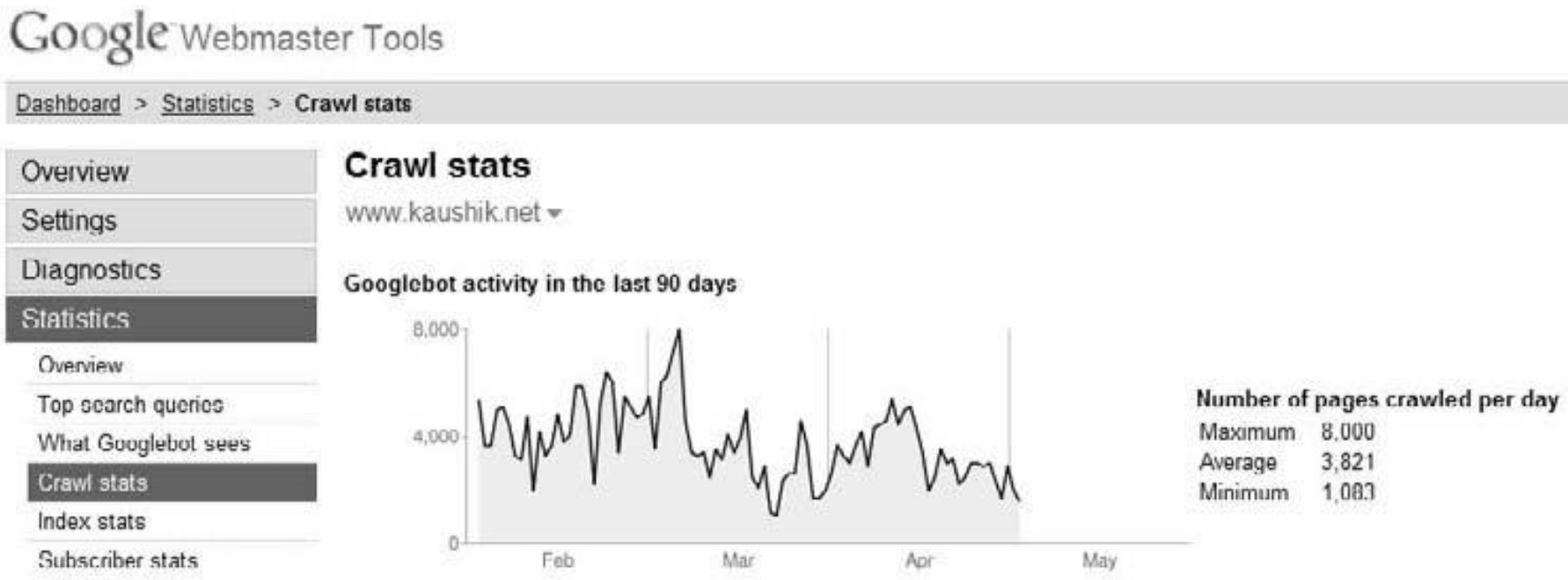
Kaushik, A. (2009). *Web analytics 2.0: The art of online accountability and science of customer centricity*. John Wiley & Sons.

Coverage

- You can measure the impact of your SEO efforts in terms of content coverage in two ways:
  - You can measure the amount of content being indexed over time (this should go up if you publish new content)
  - You can measure the number of pages on your site that get traffic from search engines

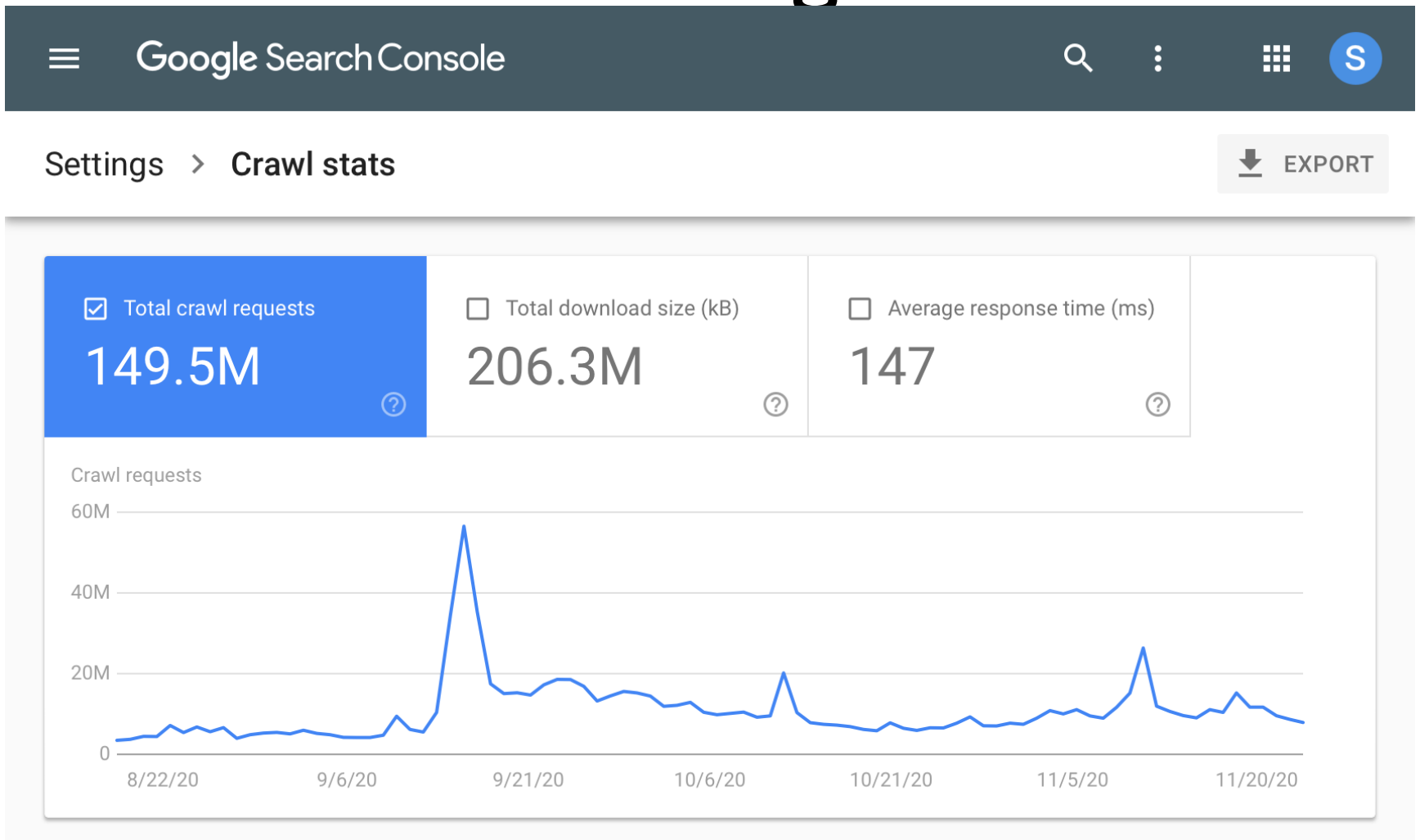
Kaushik, A. (2009). *Web analytics 2.0: The art of online accountability and science of customer centricity*. John Wiley & Sons.

Coverage



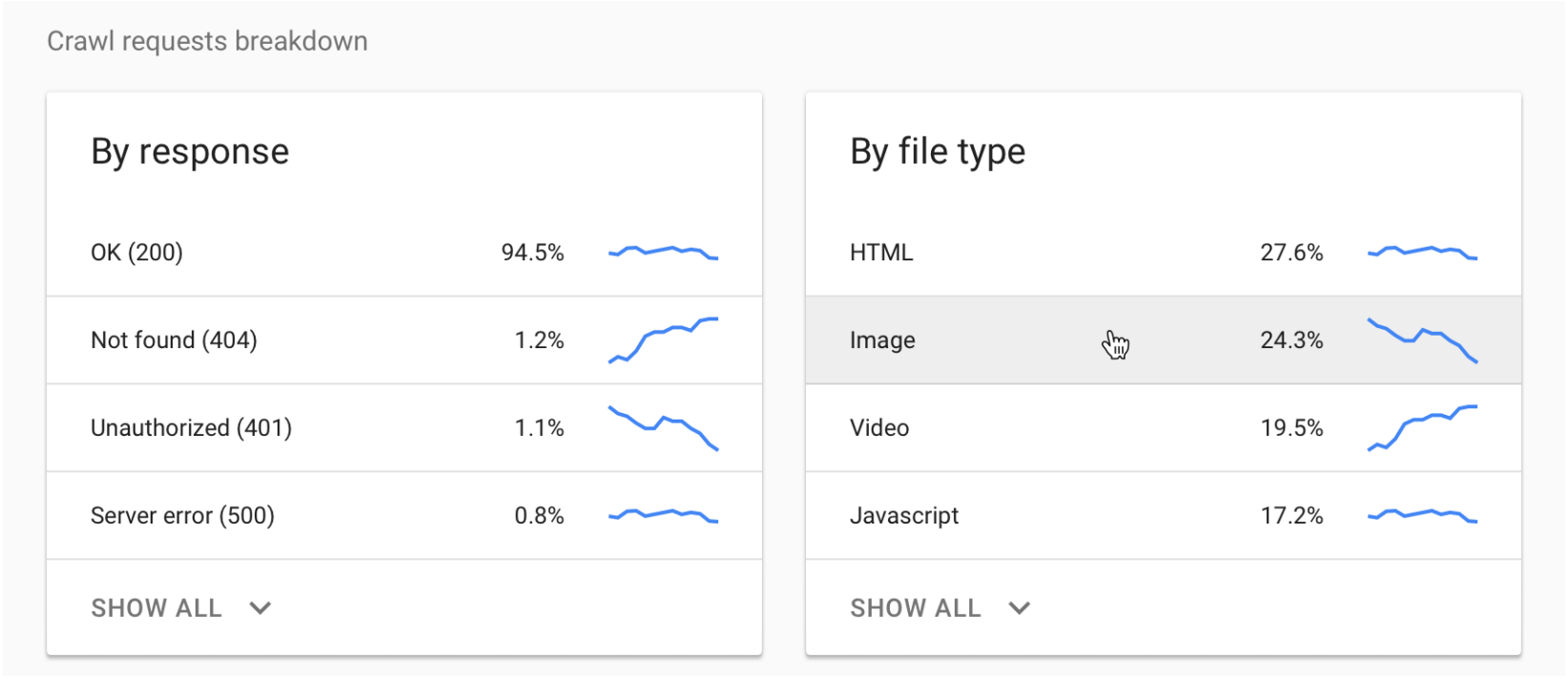
Kaushik, A. (2009). *Web analytics 2.0: The art of online accountability and science of customer centricity*. John Wiley & Sons.

Coverage

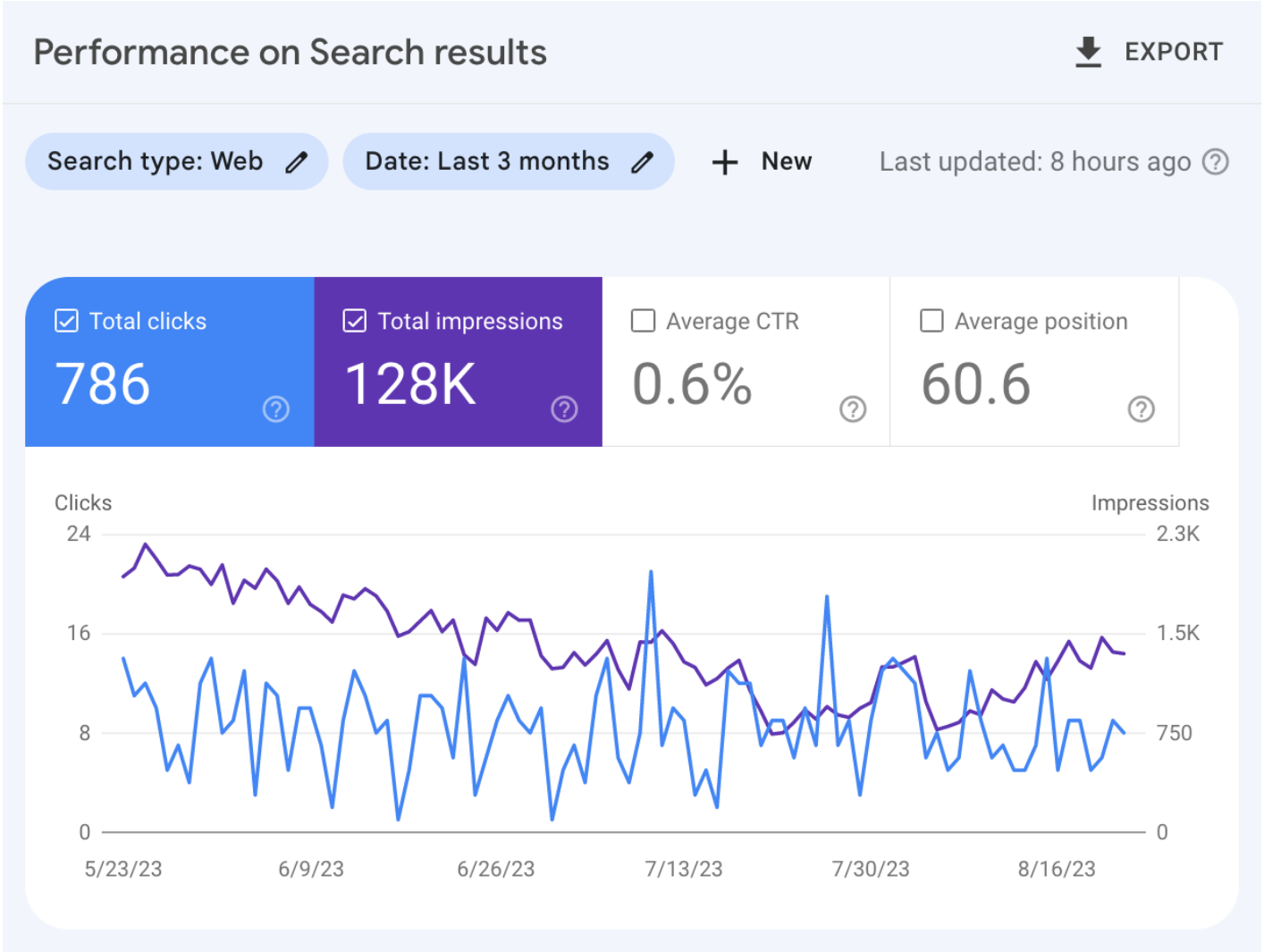


Google Dev Guide

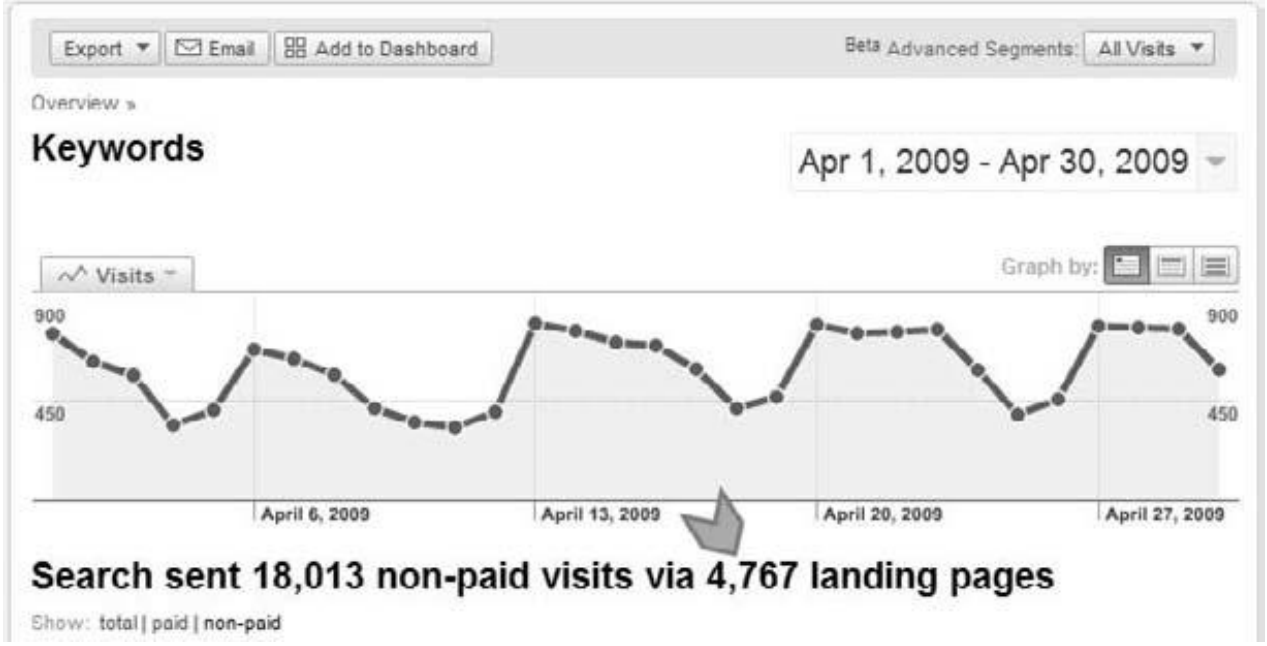
Coverage


Crawl requests breakdown
By response
OK (200) 94.5%
Not found (404) 1.2%
Unauthorized (401) 1.1%
Server error (500) 0.8%
SHOW ALL
By file type
HTML 27.6%
Image 24.3%
Video 19.5%
Javascript 17.2%
SHOW ALL

Coverage



Google Dev Guide



Kaushik, A. (2009). *Web analytics 2.0: The art of online accountability and science of customer centricity*. John Wiley & Sons.

Keyword Performance

- Need to look beyond the keywords in your analytics reports and try to understand whether you are showing up for the optimal set of keywords in search engine results
- An excellent way to measure these surprise keywords is to use the Google Search Console

Kaushik, A. (2009). *Web analytics 2.0: The art of online accountability and science of customer centricity*. John Wiley & Sons.

Top search queries

www.kaushik.net/avinash

How do I use this data?

1 week ago All searches - All locations

Top search queries
The top 20 queries in which your site appeared, and the percentage of the top 20 queries represented by each search.

#	%	Query	Position
1	29%	google analytic	7
2	11%	survey questions	6
3	8%	bounce rate	6
4	6%	google analytics	40
5	6%	analytics google	8
6	5%	competitive intelligence	8
7	4%	embarrass	10
8	4%	king baby	7
9	3%	working at google	6
10	3%	the bounce	5
11	3%	working for google	8
12	3%	statistically significant	7
13	2%	zaaz	3
14	2%	google anlytics	4
15	2%	coradiant	5
16	2%	4q	5
17	2%	avinash kaushik	2
18	2%	avinash	2
19	1%	work at google	9
20	1%	"buy in"	5

Top clicked queries
The top 20 queries from which users reached your site, and the percentage of the top 20 queries represented by each click.

#	%	Query	Position
1	42%	hippo	8
2	9%	survey questions	6
3	5%	water drop	17
4	4%	water drops	26
5	4%	trinity	12
6	4%	enterprise	17
7	3%	choice	8
8	3%	working at google	6
9	3%	drop of water	9
10	3%	unique	6
11	3%	evolution timeline	3
12	3%	tear drops	7
13	2%	variables	7
14	2%	to be or not to be	6
15	2%	google analytics	2
16	2%	liberty of the seas	17
17	2%	problems	4
18	2%	competitive intelligence	8
19	1%	avinash kaushik	2
20	1%	ripple	34

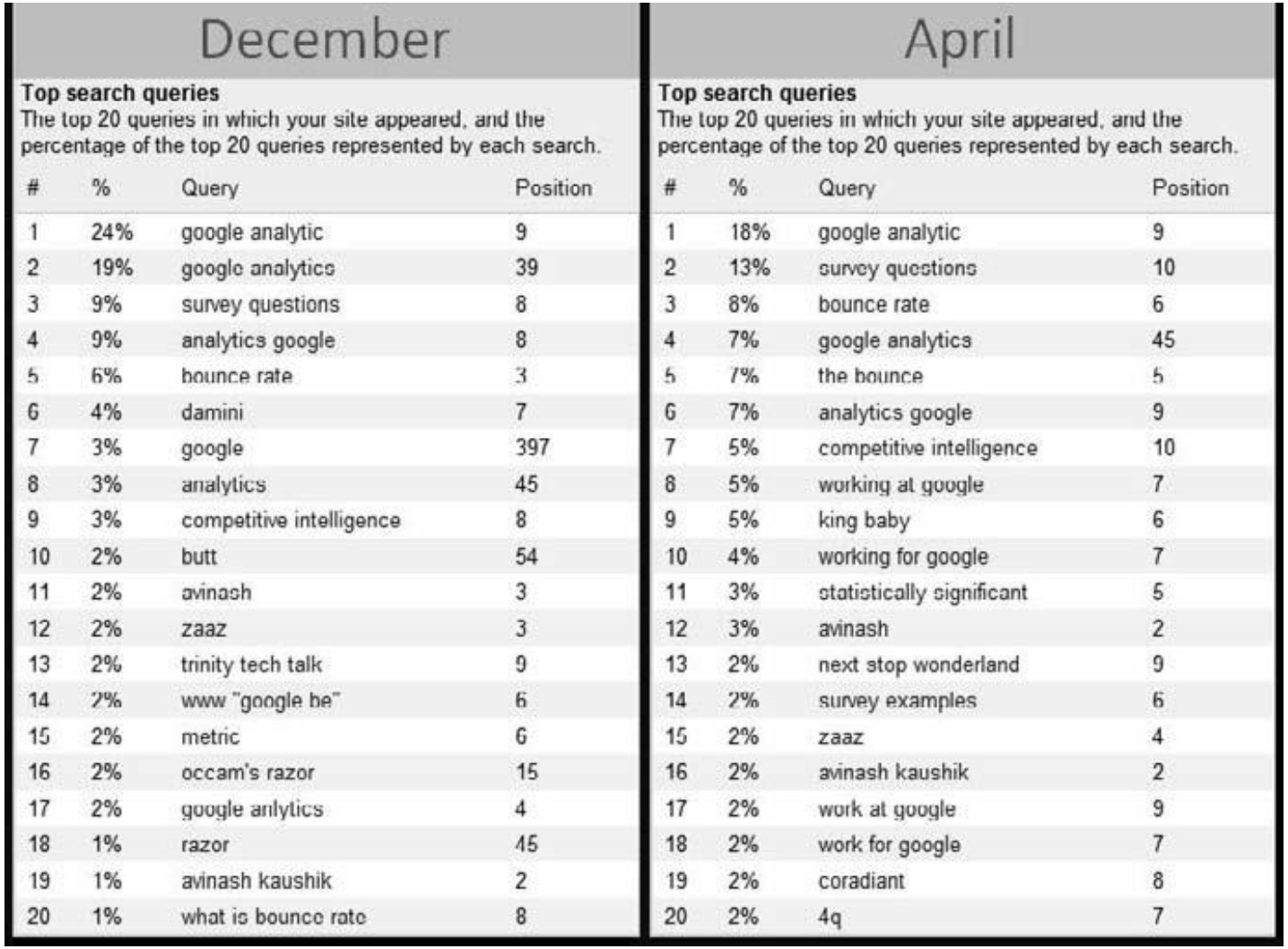
Download data Download all query stats for this site (including subfolders)

Kaushik, A. (2009). *Web analytics 2.0: The art of online accountability and science of customer centricity*. John Wiley & Sons.

Keyword Performance

- When you analyze the data, look for surprises
- Your site shows up for relevant results (1, 2, 3, 4) and less than relevant results (7, 8, 20)
- You can optimize your **robots.txt** file or your website content to ensure you are showing up for all the right keywords

Kaushik, A. (2009). *Web analytics 2.0: The art of online accountability and science of customer centricity*. John Wiley & Sons.



December

Top search queries
The top 20 queries in which your site appeared, and the percentage of the top 20 queries represented by each search.

#	%	Query	Position
1	24%	google analytic	9
2	19%	google analytics	39
3	9%	survey questions	8
4	9%	analytics google	8
5	6%	bounce rate	3
6	4%	damini	7
7	3%	google	397
8	3%	analytics	45
9	3%	competitive intelligence	8
10	2%	butt	54
11	2%	avinash	3
12	2%	zaaz	3
13	2%	trinity tech talk	9
14	2%	www "google be"	6
15	2%	metric	6
16	2%	occam's razor	15
17	2%	google anlytics	4
18	1%	razor	45
19	1%	avinash kaushik	2
20	1%	what is bounce rate	8

April

Top search queries
The top 20 queries in which your site appeared, and the percentage of the top 20 queries represented by each search.

#	%	Query	Position
1	18%	google analytic	9
2	13%	survey questions	10
3	8%	bounce rate	6
4	7%	google analytics	45
5	7%	the bounce	5
6	7%	analytics google	9
7	5%	competitive intelligence	10
8	5%	working at google	7
9	5%	king baby	6
10	4%	working for google	7
11	3%	statistically significant	5
12	3%	avinash	2
13	2%	next stop wonderland	9
14	2%	survey examples	6
15	2%	zaaz	4
16	2%	avinash kaushik	2
17	2%	work at google	9
18	2%	work for google	7
19	2%	coradiant	8
20	2%	4q	7

Kaushik, A. (2009). *Web analytics 2.0: The art of online accountability and science of customer centricity*. John Wiley & Sons.

Testing and Experiments

- Chemistry, biology, etc. offline tradition of testing and experimentation
- Why do not we do online?
- Testing online is not optional but ‘mandatory’ for every business of any size
- Without leveraging some components of testing, your business will never scale the magnificent heights it deserves

Kaushik, A. (2009). *Web analytics 2.0: The art of online accountability and science of customer centricity*. John Wiley & Sons.

Benefits

- You no longer have to guess what will work on your website
- You can launch early and often and fail faster; your only limitation is the amount of traffic you get and the ideas you have
- You control what success is—you identify the goal, and you decide what outcome you desire
- You allow your customers to participate as they would normally experience your site and help you pick winners (all without knowing that they are in a test!)

Kaushik, A. (2009). *Web analytics 2.0: The art of online accountability and science of customer centricity*. John Wiley & Sons.

Benefits

- You can experiment for free if cost is a barrier, using tools such as the free Google Optimize, and you can start tomorrow (no RFP, no complex product evaluation, no company bureaucracy)
- You don’t have to buy a web analytics tool or worry about reporting or building your own multivariate or regression platforms for content or testing results

Kaushik, A. (2009). *Web analytics 2.0: The art of online accountability and science of customer centricity*. John Wiley & Sons.

Testing Options

- A/B testing
- Multivariate Testing (MVT)

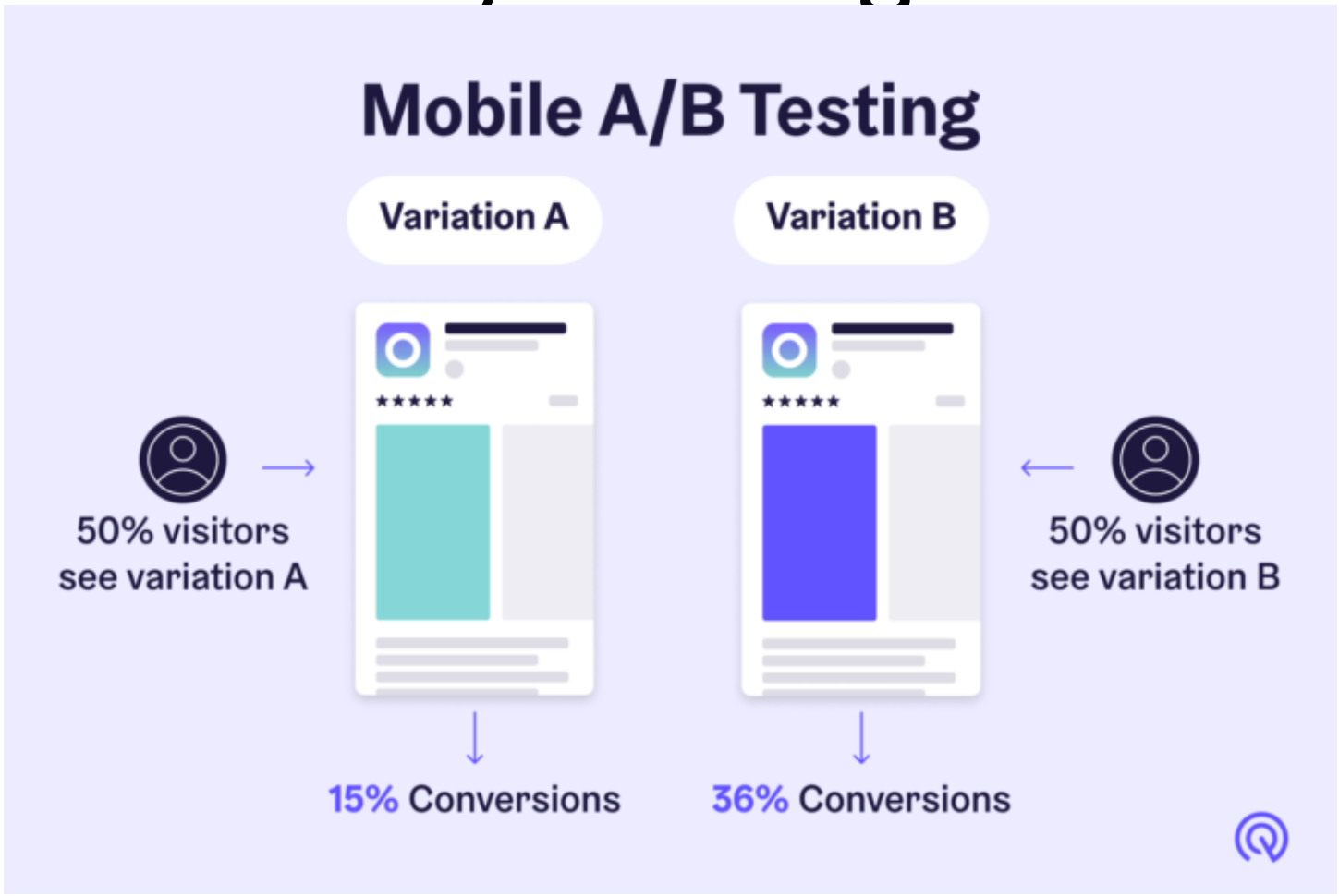
Kaushik, A. (2009). *Web analytics 2.0: The art of online accountability and science of customer centricity*. John Wiley & Sons.

A/B Testing

- A/B testing is a technique for testing two or more versions of a page on your website
- Each version of the web page is unique and can be visually differentiated from the control (the original page) without too much effort

Kaushik, A. (2009). *Web analytics 2.0: The art of online accountability and science of customer centricity*. John Wiley & Sons.

A/B Testing


Mobile A/B Testing
Variation A
Variation B
50% visitors
see variation A
50% visitors
see variation B
15% Conversions
36% Conversions

A/B Testing

- The goal is to try a few versions of the page and identify which version delivers the desired outcome (for example, more click-throughs, more Conversions, lower Bounce Rates, etc.)
- To accomplish the test, each version is randomly shown to a predetermined percentage of people who come to the page, usually 50 percent (or in the case of an A/B/C test: 33 percent, 33 percent, 34 percent)

Kaushik, A. (2009). *Web analytics 2.0: The art of online accountability and science of customer centricity*. John Wiley & Sons.

A/B Testing

- A/B tests are best at testing big changes to the layout and templates of your web pages or for cases where you want to add or remove pages from a normal structured experience
- for example, it takes four steps to apply for a credit card—you can test two steps or five steps

Kaushik, A. (2009). *Web analytics 2.0: The art of online accountability and science of customer centricity*. John Wiley & Sons.

A/B Testing

- Getting started with A/B tests is pretty straightforward
- Start by looking at the Top Landing Pages report, and identify the pages with the highest Bounce Rates
- Then have a brainstorming exercise for ideas to improve the page, coalesce the most promising, and get (beg) your designers and developers to create one or two versions with the new ideas
- Launch the new ideas and measure

Kaushik, A. (2009). *Web analytics 2.0: The art of online accountability and science of customer centricity*. John Wiley & Sons.

Pros of A/B Testing

- A/B testing is probably the cheapest way to start testing because you'll use existing resources in your company, the tools are free, or you can even use your website platform
- A/B testing is not very difficult in terms of effort
- Without a doubt, A/B testing is the best technique to start your testing journey
- The results are easy to communicate

Kaushik, A. (2009). *Web analytics 2.0: The art of online accountability and science of customer centricity*. John Wiley & Sons.

Cons of A/B Testing

- A/B testing is an all-or-nothing approach
- You may know which page won quite easily, but it may be harder to tease out which elements contributed the most or not at all
- Too much reliance on A/B could mean you are making changes too slowly

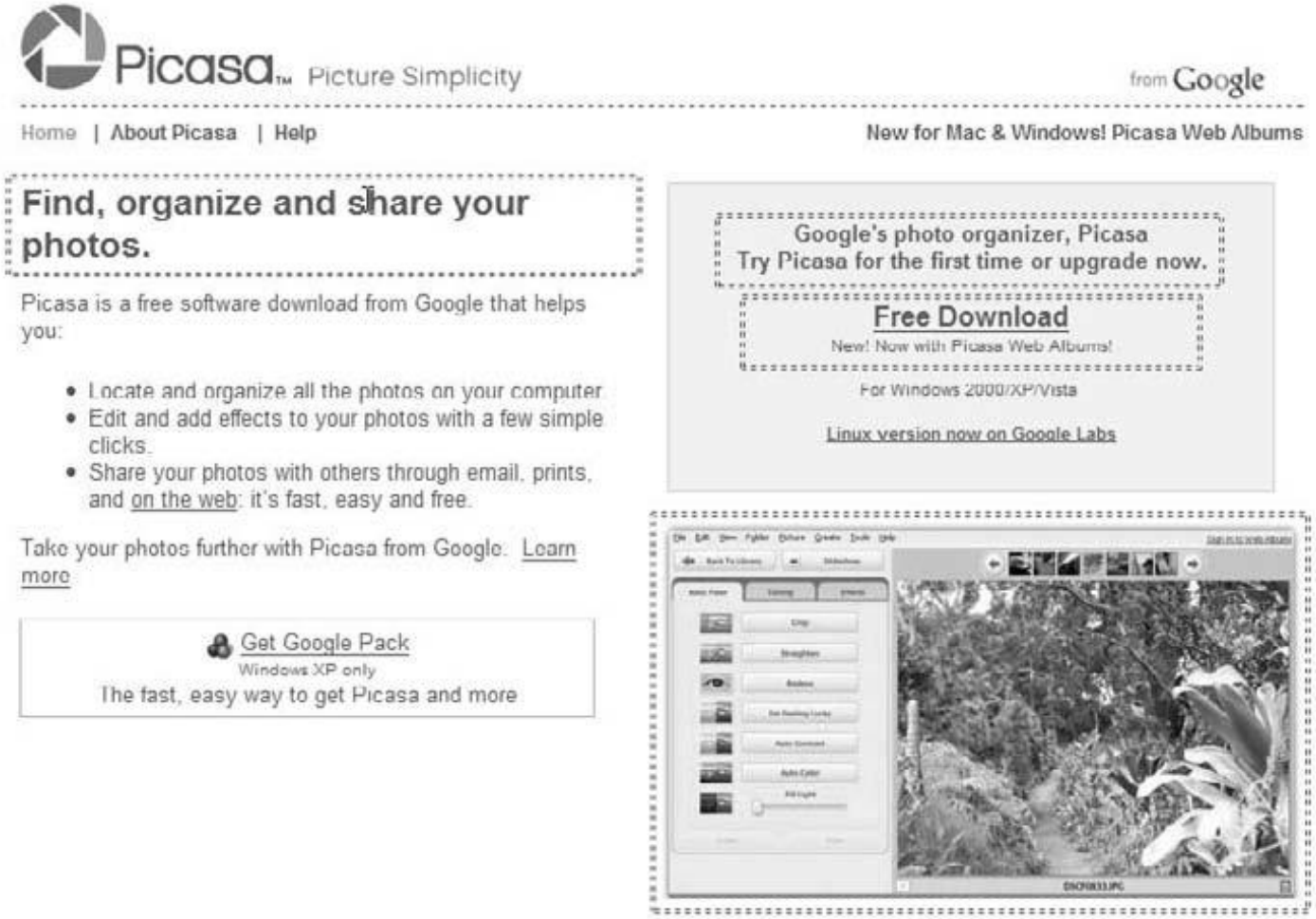
Kaushik, A. (2009). *Web analytics 2.0: The art of online accountability and science of customer centricity*. John Wiley & Sons.

Multivariate Testing

- Multivariate testing is a technique for testing changes to many different elements all at the same time on one web page

Kaushik, A. (2009). *Web analytics 2.0: The art of online accountability and science of customer centricity*. John Wiley & Sons.

Multivariate Testing



Kaushik, A. (2009). *Web analytics 2.0: The art of online accountability and science of customer centricity*. John Wiley & Sons.

Multivariate Testing

- Four elements, identified in Figure with double-dotted lines, are critical on the page and will be optimized in this test
- For example, you would test four different versions of the main text on this page: “Find, organize and share your photos” and a few versions of the main image (bottom right)
- You would also test the page with no image at all and with three alternative calls to action instead of “Free Download”

Kaushik, A. (2009). *Web analytics 2.0: The art of online accountability and science of customer centricity*. John Wiley & Sons.

Multivariate Testing (Technical Side)

- To execute MVT, you would implement the JavaScript tags of the testing tool on the page
- In the previous example, you would place tags around the four identified elements and upload the alternatives into the testing tool
- Testing tool examples – Optimizely, SiteSpect, VWO

Kaushik, A. (2009). *Web analytics 2.0: The art of online accountability and science of customer centricity*. John Wiley & Sons.

Multivariate Testing

- Based on the settings you choose, visitors to the page are shown a dynamically created page with a variation of the elements in the test—in this case a different image, call to action, and main body text
- Your MVT tool will start to produce the data almost immediately

Kaushik, A. (2009). *Web analytics 2.0: The art of online accountability and science of customer centricity*. John Wiley & Sons.

Multivariate Testing

Combinations | **Page Sections**

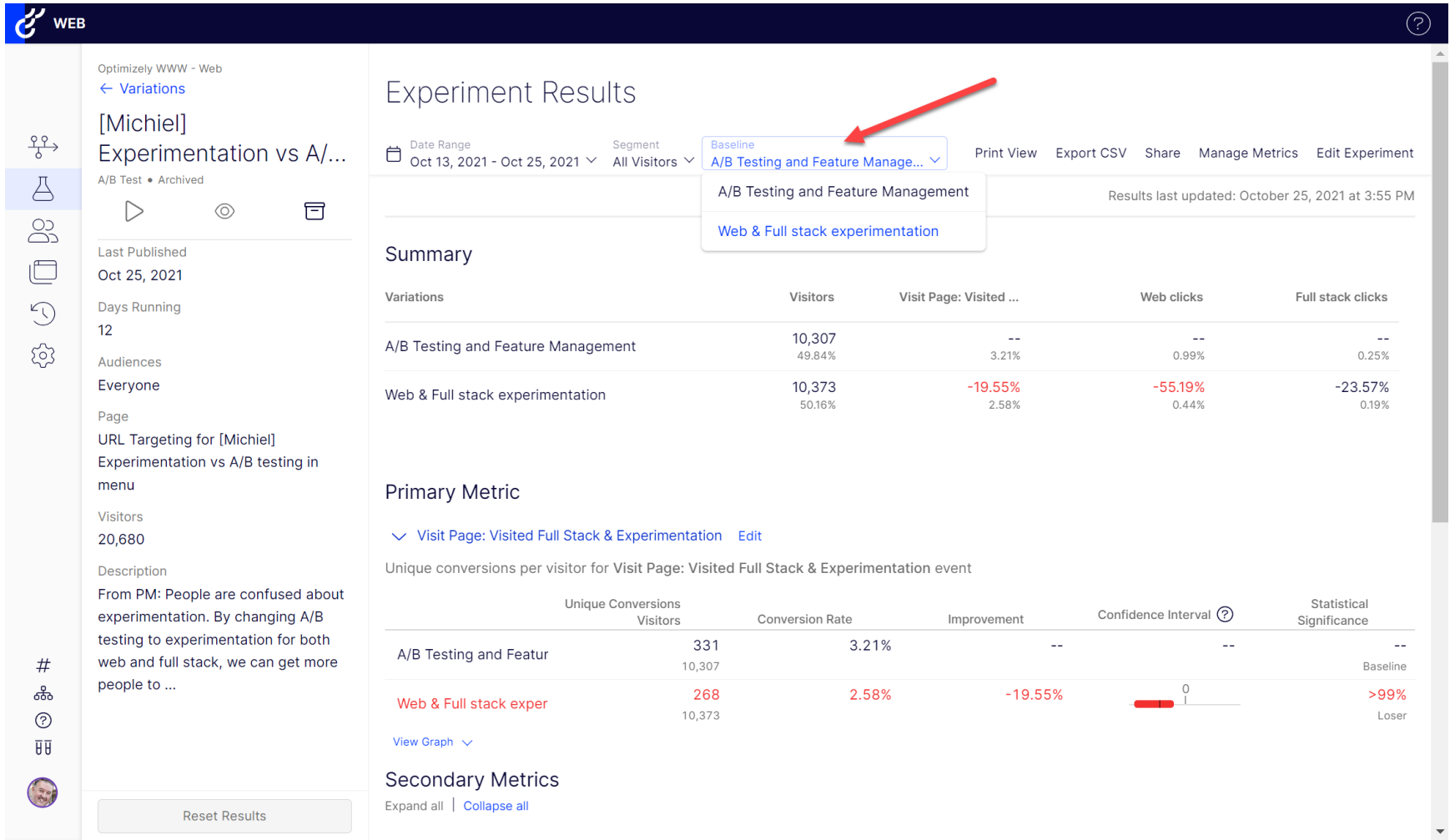
Analysis

Sort By: (•) Relevance Rating ( ) Order Created **Download:** T | Print | Preview

Relevance Rating [?]	Variation	Estimated Conversion Rate Range [?]	Chance to Beat Orig. [?]	Chance to Beat All [?]	Observed Improvement [?]	Conversions / Impressions [?]
Section 3 3 / 5	Original	**50.1%** ± 0.4%	—	4.97%	—	12456 / 24859
	Variation 3	**50.8%** ± 0.4%	94.4%	91.0%	1.42%	12693 / 24978
	Variation 1	**50.1%** ± 0.4%	46.9%	4.02%	-0.07%	12539 / 25042
	Variation 2	**49.4%** ± 0.4%	5.18%	0.03%	-1.46%	12404 / 25121
Section 2 2 / 5	Original	**50.2%** ± 0.3%	—	14.0%	—	16818 / 33532
	Variation 1	**50.6%** ± 0.3%	86.0%	85.9%	0.82%	16936 / 33491
	Variation 2	**49.5%** ± 0.4%	5.65%	0.17%	-1.22%	16338 / 32977
Section 1 1 / 5	Original	**50.3%** ± 0.3%	—	94.6%	—	24944 / 49544
	Variation 1	**49.8%** ± 0.3%	5.37%	5.37%	-1.00%	25148 / 50456
Section 4 0 / 5	Original	**50.1%** ± 0.5%	—	10.0%	—	10033 / 20035
	Variation 3	**50.5%** ± 0.5%	79.8%	50.6%	0.83%	10109 / 20020
	Variation 2	**50.3%** ± 0.5%	66.1%	24.2%	0.41%	9964 / 19815
	Variation 1	**50.2%** ± 0.5%	57.3%	14.9%	0.18%	9994 / 19921
	Variation 4	**49.4%** ± 0.5%	10.0%	0.21%	-1.27%	9992 / 20209

Kaushik, A. (2009). *Web analytics 2.0: The art of online accountability and science of customer centricity*. John Wiley & Sons.

Multivariate Testing



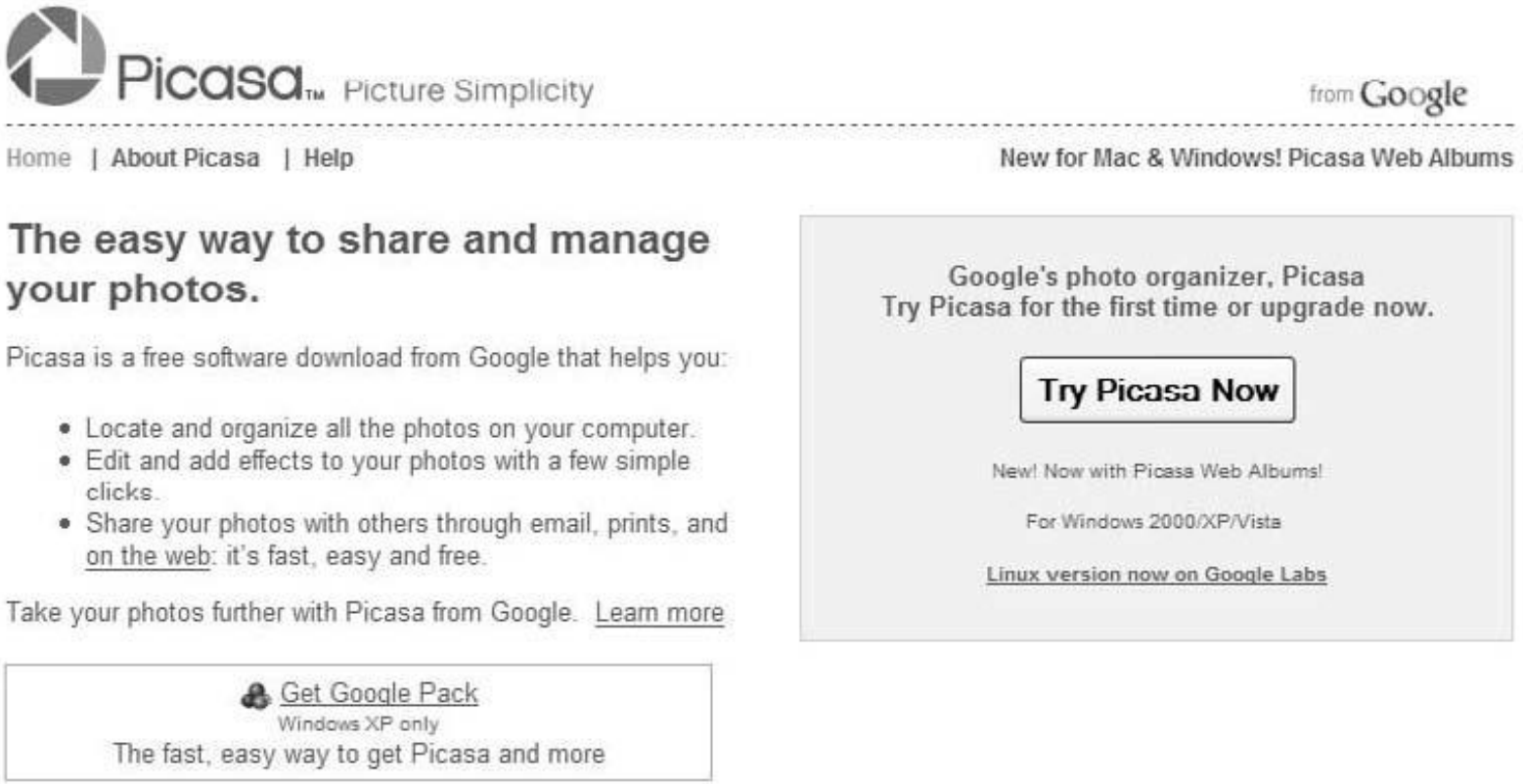
Optimizely Dev Guide: https://support.optimizely.com/hc/en-us/articles/4410284017421-Optimizely-Experimentation-Results-page

Multivariate Testing

- In the case of Google Website Optimizer, you are keeping a close eye on the Chance to Beat Orig. column
- Your focus will be on the Observed Improvement column

Kaushik, A. (2009). *Web analytics 2.0: The art of online accountability and science of customer centricity*. John Wiley & Sons.

Testing Analysis



Kaushik, A. (2009). *Web analytics 2.0: The art of online accountability and science of customer centricity*. John Wiley & Sons.

Testing Analysis

- It is not a huge surprise that the text “The easy way to share and manage your photos” is better than “Find, organize and share your photos”
- It is surprising, however, that the image showing the program benefits did not help Conversions!
- Simple text worked better than buttons, especially with the word Free, but in this case the “Free Download” text link did not work as well as the button “Try Picassa Now”
- Would you ever guess that “Try” works better than “Free”? We can find out with experiments

Kaushik, A. (2009). *Web analytics 2.0: The art of online accountability and science of customer centricity*. John Wiley & Sons.

Common MVT Techniques

- Two techniques are commonly used in executing multivariate tests:
- Full factorial
- Partial factorial (also known as fractional factorial)

Kaushik, A. (2009). *Web analytics 2.0: The art of online accountability and science of customer centricity*. John Wiley & Sons.

Common MVT Techniques

- With full factorial tests, you test all combinations of the pages that might occur as a result of your experiment
- For example, in the previous test, you have four variations of all four elements—image, text, call to action, copy—so you have 4 × 4 × 4 × 4 = 256 combinations

Kaushik, A. (2009). *Web analytics 2.0: The art of online accountability and science of customer centricity*. John Wiley & Sons.

Common MVT Techniques

- With partial factorial tests, you test fewer combinations and infer results for what might happen with other combinations
- A full factorial test is great at teasing out key data around interactions between each element, and it helps you make a more informed decision
- But full factorial comes at a cost: the time required and the number of test participants (as you need enough data for each combination)

Kaushik, A. (2009). *Web analytics 2.0: The art of online accountability and science of customer centricity*. John Wiley & Sons.

Common MVT Techniques

- A partial factorial test loses some richness of data, but you can get results faster
- Another hybrid option is gaining acceptance; in this method you would use partial factorial to identify the worst combinations, prune them, and run a full factorial test for remaining promising combinations

Kaushik, A. (2009). *Web analytics 2.0: The art of online accountability and science of customer centricity*. John Wiley & Sons.

Pros of MVT Testing

- Doing multivariate testing gets easier with each passing day
- You can make things as complicated as you want, but you can also start very simply with just a few lines of JavaScript tags on a page
- Almost every aspect of MVT is outsourced – SaaS
- The amount of data collected with MVT directly translates into an enhanced understanding of your visitor preferences and the value of your ideas, directly translates into a more optimal website experience

Kaushik, A. (2009). *Web analytics 2.0: The art of online accountability and science of customer centricity*. John Wiley & Sons.

Cons of MVT Testing

- MVT does require more effort and a higher level of commitment from your organization
- Although the rewards are rich, you are also putting a lot of resources into the process
- Your experiments are only as good as the ideas you put into them, more research
- Your website experience is not a single page

Kaushik, A. (2009). *Web analytics 2.0: The art of online accountability and science of customer centricity*. John Wiley & Sons.

Regression Analysis

- Regression analysis is used to investigate the relationship between two or more variables
- Often, we are interested in predicting a variable using one or more independent variables x1, x2 ,..., xk
- For example, we might be interested in the relationship between two variables: sales and profit for a chain of stores, number of hours required to produce a certain number of products, number of accidents vs. blood alcohol level, advertising expenditures and sales, or the height of parents compared to their children

Sahay, A. (2018). *Business Analytics, Volume I: A Data-Driven Decision Making Approach for Business*. Business Expert Press.

Regression Analysis

- In general, we have one dependent or response variable, y and one or more independent variables, x1, x2 ,…., xk
- The independent variables are also called predictors
- If there is only one independent variable x that we are trying to relate to the dependent variable y, then this is a case of simple regression
- On the other hand, if we have two or more independent variables that are related to a single response or dependent variable, then we have a case of multiple regression

Sahay, A. (2018). *Business Analytics, Volume I: A Data-Driven Decision Making Approach for Business*. Business Expert Press.

Regression Analysis

- In general, we have one dependent or response variable, y and one or more independent variables, x1, x2 ,…., xk
- The independent variables are also called predictors
- If there is only one independent variable x that we are trying to relate to the dependent variable y, then this is a case of simple regression
- On the other hand, if we have two or more independent variables that are related to a single response or dependent variable, then we have a case of multiple regression

Sahay, A. (2018). *Business Analytics, Volume I: A Data-Driven Decision Making Approach for Business*. Business Expert Press.

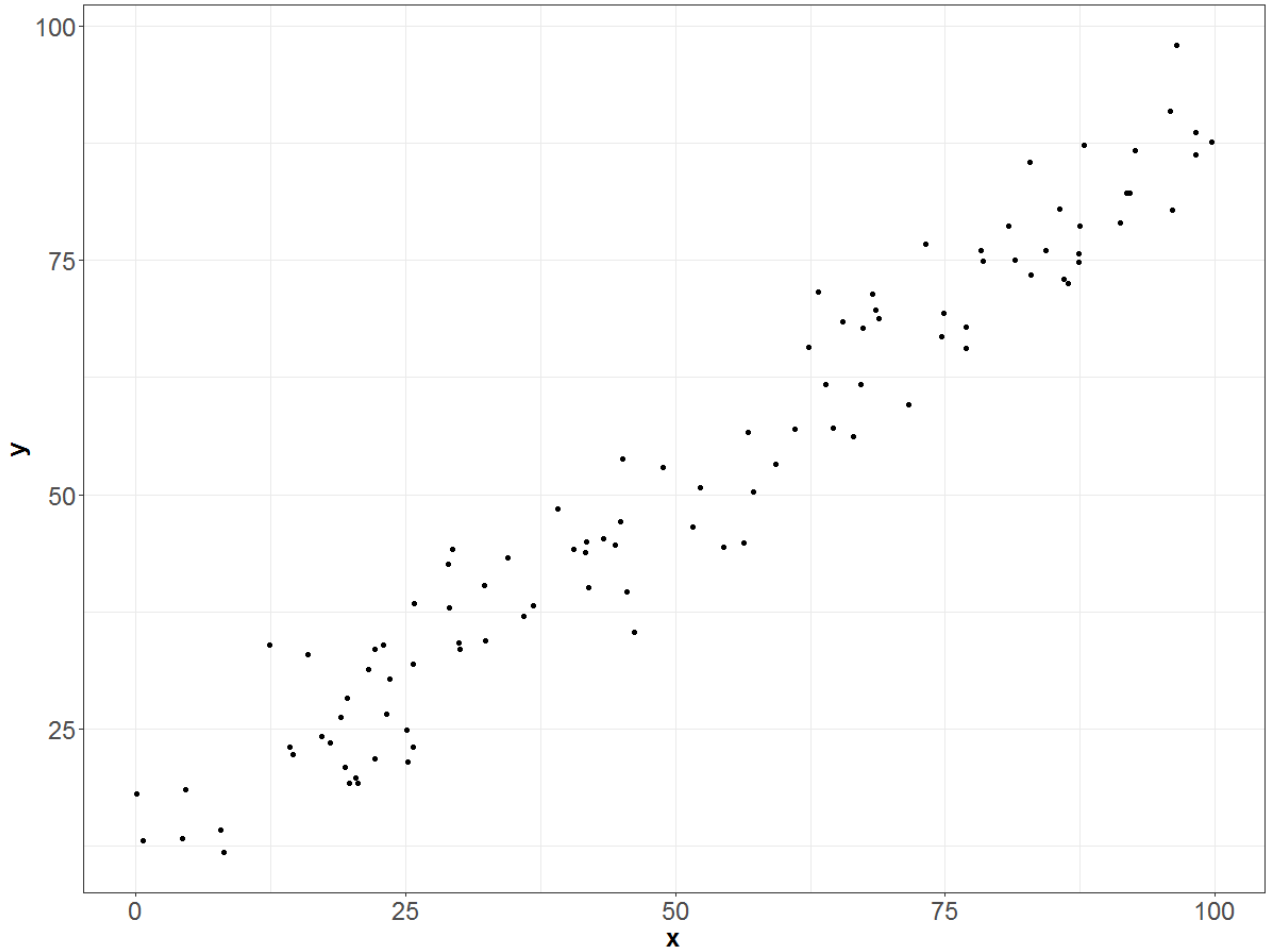
Regression Analysis

- In regression analysis, the dependent or response variable y is a random variable; whereas the independent variable or variables x1, x2 ,..., xn are measured with negligible error and are controlled by the analyst

- The relationship between the dependent and independent variable or variables are described by a mathematical model known as a regression model

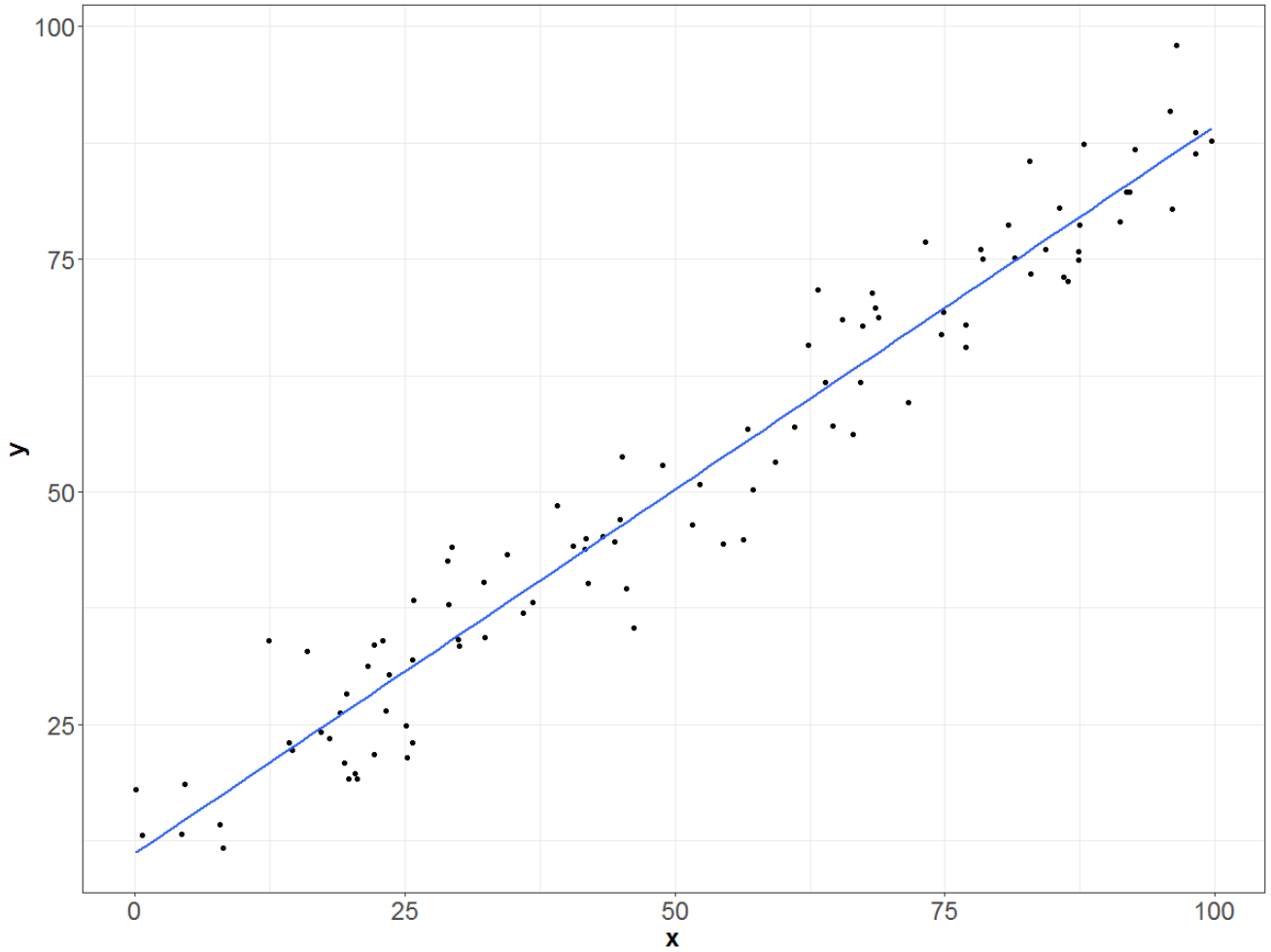
Sahay, A. (2018). *Business Analytics, Volume I: A Data-Driven Decision Making Approach for Business*. Business Expert Press.

Regression Analysis



Sahay, A. (2018). *Business Analytics, Volume I: A Data-Driven Decision Making Approach for Business*. Business Expert Press.

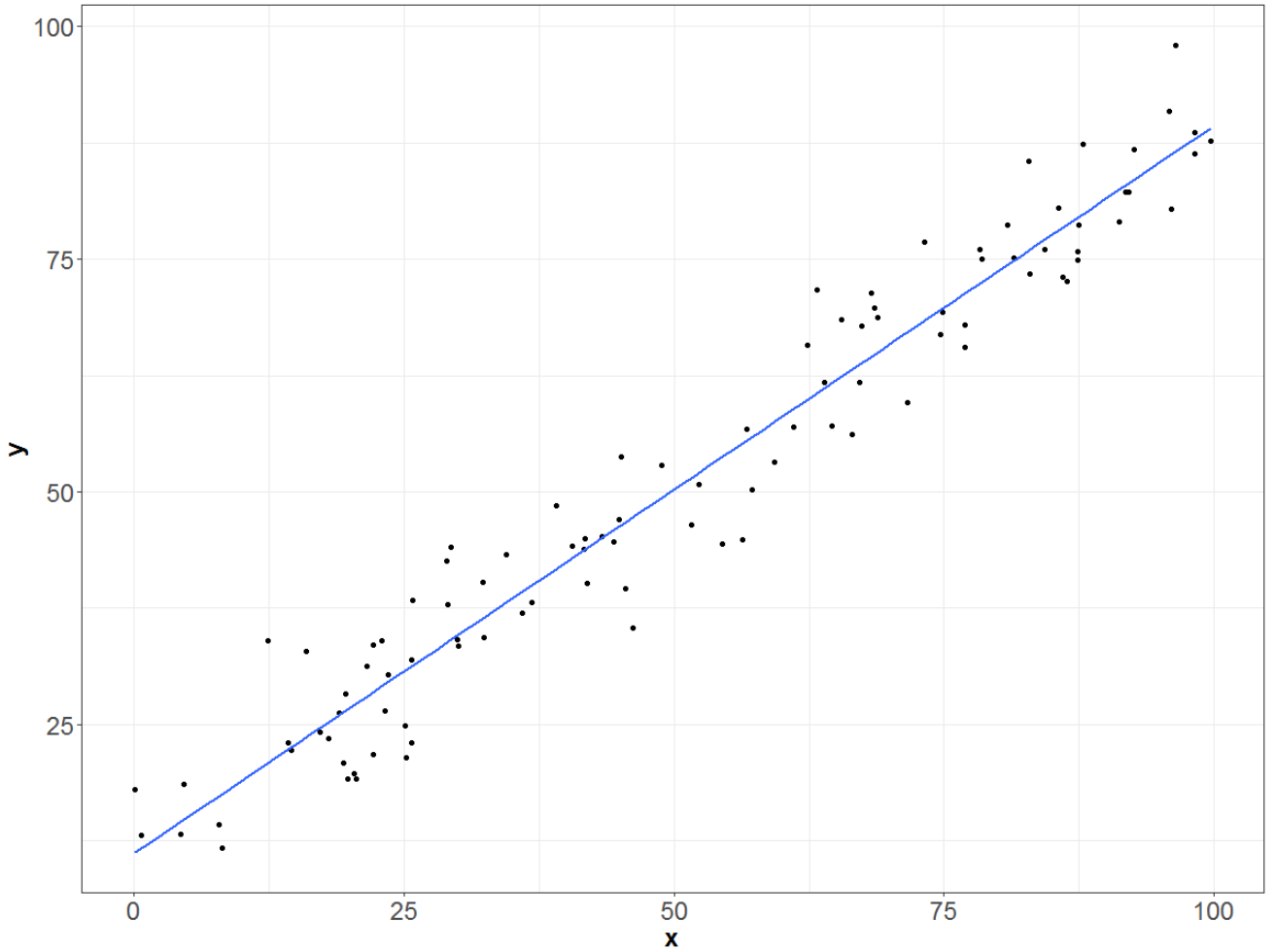
Regression Analysis



Sahay, A. (2018). *Business Analytics, Volume I: A Data-Driven Decision Making Approach for Business*. Business Expert Press.

Regression Analysis

- $y=b+mx$
- b is the y-intercept
- m is the slope



Sahay, A. (2018). *Business Analytics, Volume I: A Data-Driven Decision Making Approach for Business*. Business Expert Press.

Linear Regression

- Our goal is to use the data to estimate the “best” line through the data- specifically the intercept and the slope

- In linear regression, we typically use the symbol β for parameters (values of the slope and intercept)

Sahay, A. (2018). *Business Analytics, Volume I: A Data-Driven Decision Making Approach for Business*. Business Expert Press.

Linear Regression

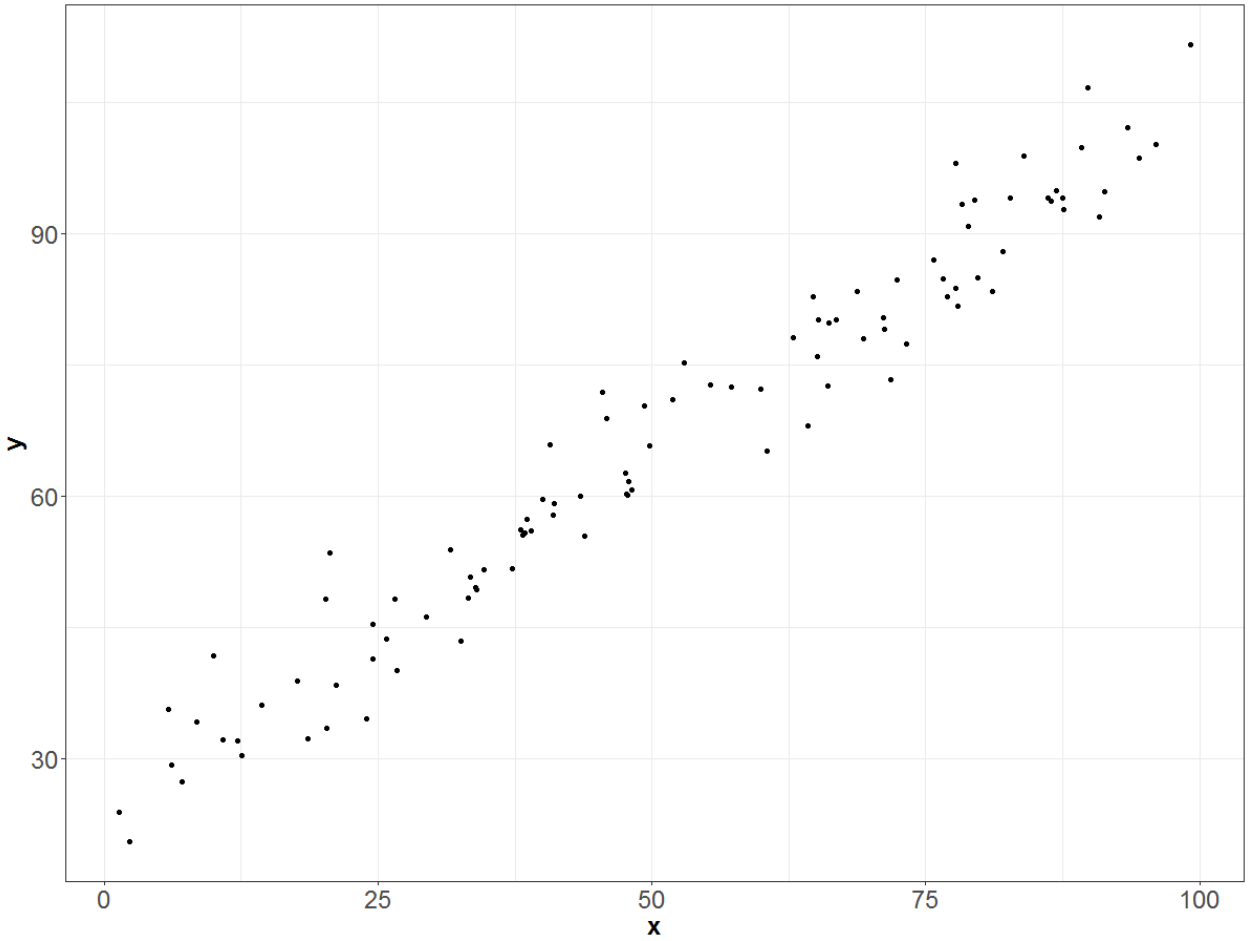
- So we try to estimate the following β0 and β1:
- Y= β_0+ β_1 X

- β_0 is the y intercept
- β_1 is the slope

Sahay, A. (2018). *Business Analytics, Volume I: A Data-Driven Decision Making Approach for Business*. Business Expert Press.

Regression Analysis

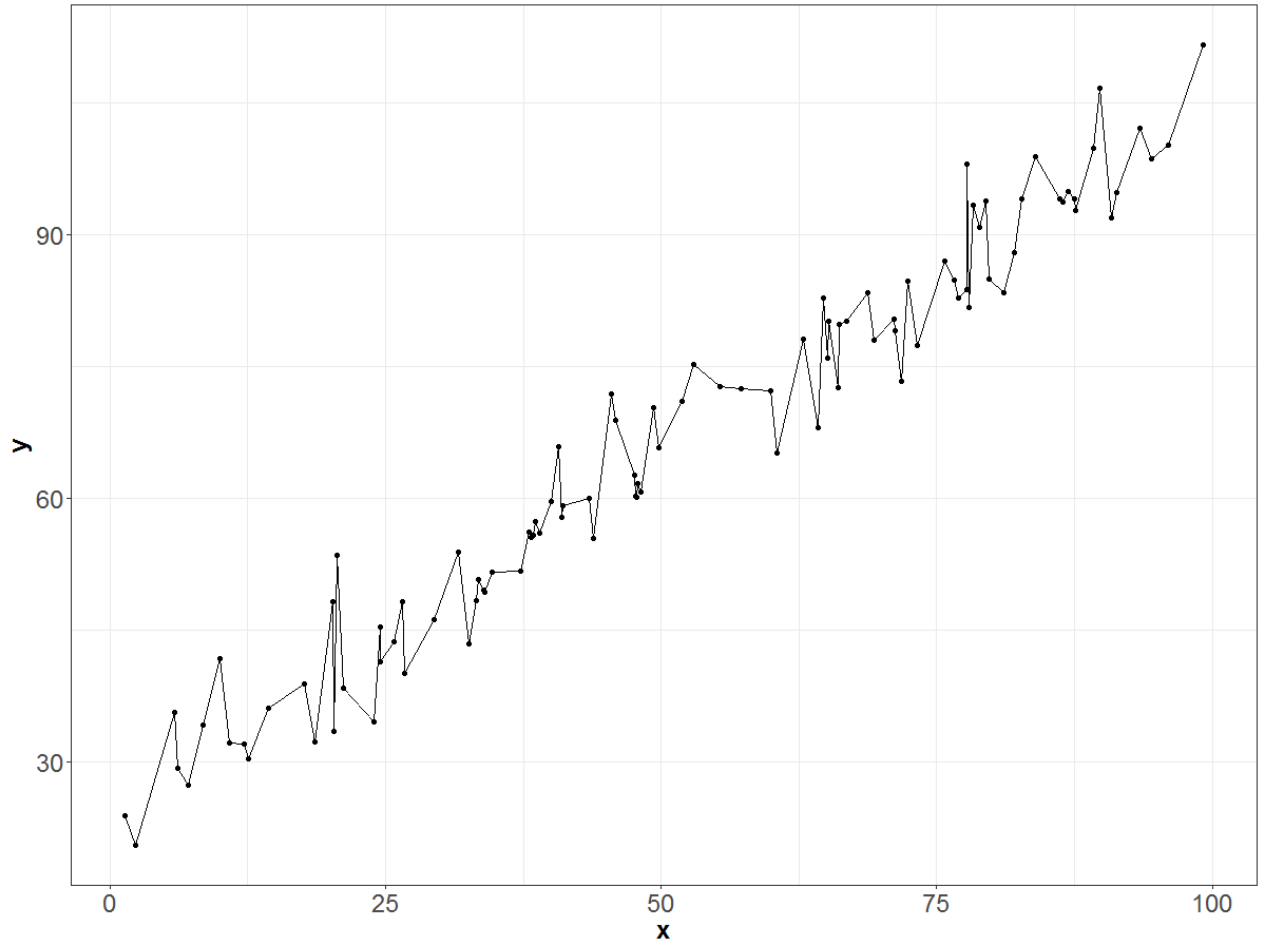
- However, we cannot find a slope and intercept that perfectly fits the data
- For example, the slopes between any two observations are different



Sahay, A. (2018). *Business Analytics, Volume I: A Data-Driven Decision Making Approach for Business*. Business Expert Press.

Regression Analysis

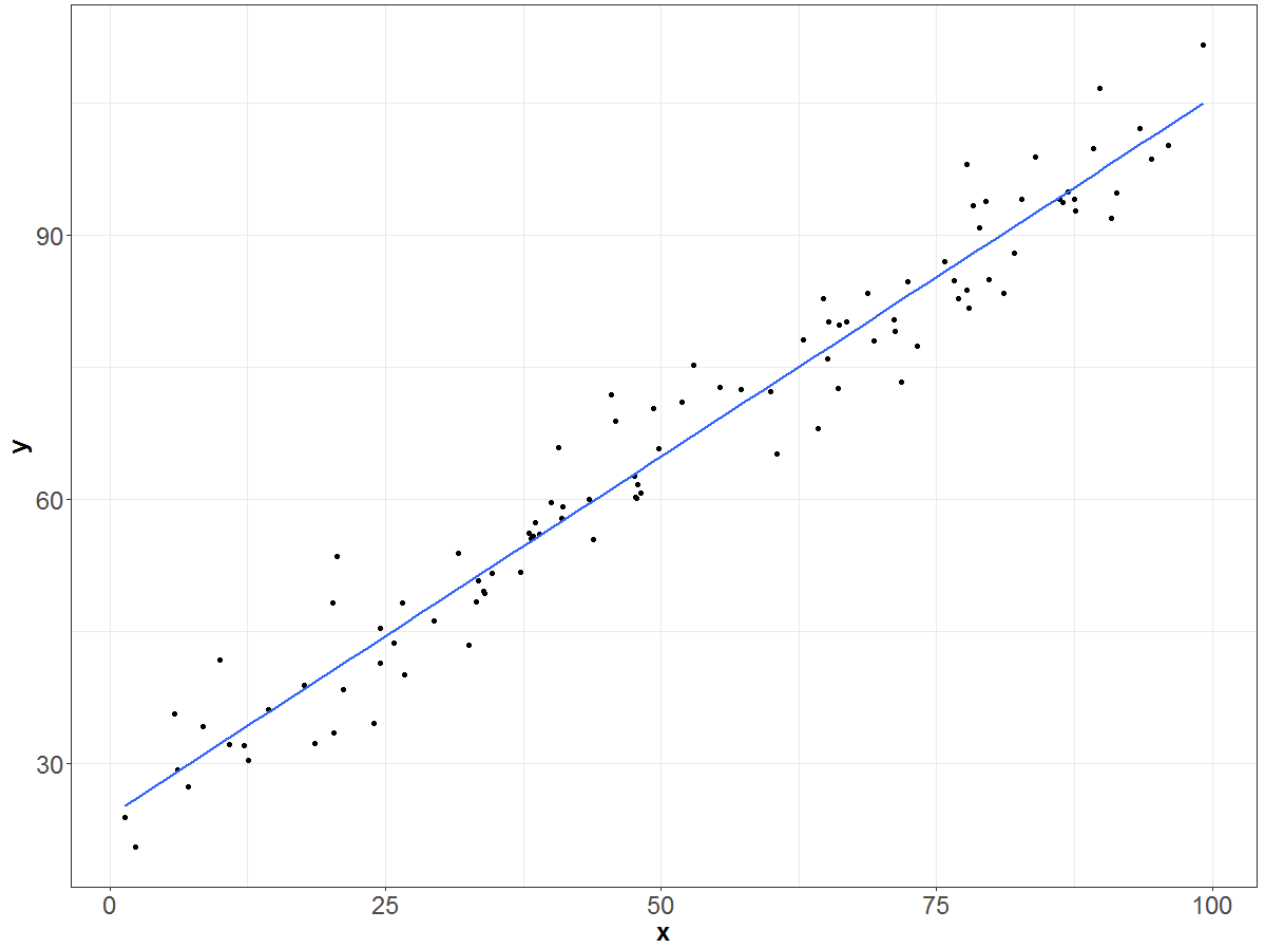
- For example – the slopes between any two observations are different



Sahay, A. (2018). *Business Analytics, Volume I: A Data-Driven Decision Making Approach for Business*. Business Expert Press.

Regression Analysis

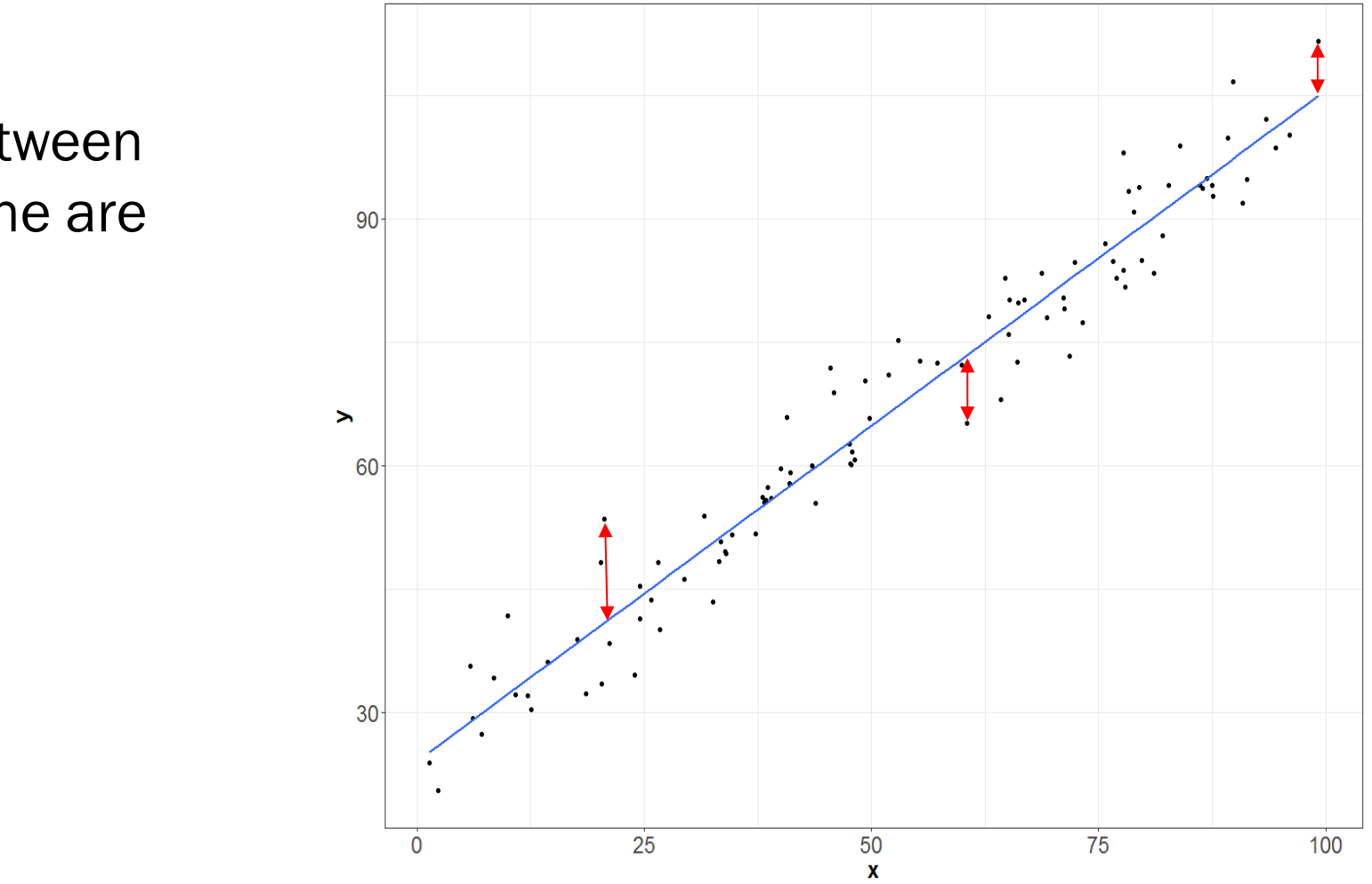
- So we do the best we can at drawing ONE line through the data…



Sahay, A. (2018). *Business Analytics, Volume I: A Data-Driven Decision Making Approach for Business*. Business Expert Press.

Regression Analysis

- The differences between the data and our line are called residuals



Sahay, A. (2018). *Business Analytics, Volume I: A Data-Driven Decision Making Approach for Business*. Business Expert Press.

Simpler Linear Regression Model

- Y_i= β_0+ β_1 X_i + ε_i
- β_0 is the y-intercept
- β_1 is the slope
- ε_i is the residual for observation i=1,2,…N
- X_i is the independent variable
- Y_i is the dependent variable

Sahay, A. (2018). *Business Analytics, Volume I: A Data-Driven Decision Making Approach for Business*. Business Expert Press.

Simpler Linear Regression Model

- Y_i= β_0+ β_1 X_i + ε_i
- β_0 is the y-intercept
- β_1 is the slope
- ε_i is the residual for observation i=1,2,…N
- X_i is the independent variable
- Y_i is the dependent variable

Sahay, A. (2018). *Business Analytics, Volume I: A Data-Driven Decision Making Approach for Business*. Business Expert Press.

Simpler Linear Regression Model

- Y_i= β_0+ β_1 X_i + ε_i
- β_0 is the y-intercept
- β_1 is the slope
- ε_i is the residual for observation i=1,2,…N
- X_i is the independent variable
- Y_i is the dependent variable

Sahay, A. (2018). *Business Analytics, Volume I: A Data-Driven Decision Making Approach for Business*. Business Expert Press.

Simpler Linear Regression Model

- Which Variable do I Make Y, and Which Variable do I make X?
- You will need to carefully read the question to find key words that tell you which is the dependent variable (Y) and which is the independent variable (X)
- A variable that is being “predicted”, “explained”, “affected”, “impacted”, etc., is the dependent variable
- On the other hand, the variable that does the predicting, explaining, affecting, etc. is the independent variable

Sahay, A. (2018). *Business Analytics, Volume I: A Data-Driven Decision Making Approach for Business*. Business Expert Press.

Simpler Linear Regression Model

- Exercise – A professor wants to know how well studying predicts test scores. What is the dependent and independent variable?
- The dependent variable is test scores, the independent variable is studying

Sahay, A. (2018). *Business Analytics, Volume I: A Data-Driven Decision Making Approach for Business*. Business Expert Press.

Assignment 5 (A/B testing at Vungle)

- Questions:

1. What would you advise Jaffer regarding the performance of the new data science algorithm?

2. Which assumptions underline your analysis?

**Submit a PDF file with your analysis in APA format (at max 5 pages).

**No excel file analysis is required for this assignment.

Thank you

vatsal.maru@utdallas.edu
JSOM 4.414