Clinical statistics for non-statisticians: Day one

Steve Simon

One warning

- Lots of real world analogies, but
  - May be too specific to U.S.A.
  - Please ask about anything obscure

Speaker notes

Speaker notes

Let me start off with a brief warning. I like to draw analogies a lot in my talks to various cultural references, such as books, television shows, or movies. I do this because it enlivens what sometimes can be a tedious topic. But I have to apologize if some of my cultural references may be too specific to the United States. Let me give an example.

There is a series, Ted Lasso, about a football coach, United States football, I mean. He is asked to coach a soccer team in England, soccer being the sport that the rest of the world calls football. Now I know that people outside the U.S.A. watch Ted Lasso. But part of the humor in that series is when Ted starts telling stories that have a point to them, and everyone stares at in befuddlement because the only people who understand that point Ted is trying to make have been living in the United States their entire lives. As an example, Ted wears a t-shirt on one show that says “Arthur Joes GAtes Stack barbecue.” It’s actually a joke that only those of us who live in Kansas City can laugh at. Look it up on the Internet if you are curious.

There’s a stereotype of people in the United States that they think the world revolves around them. It’s a stereotype that is not true for all of us, but it is true for many, including me. I try to fight that tendency, but it’s not always easy.

The point is, that I will try to be inclusive of those of you joining from outside the United States, but if I include a cultural reference that you are unfamiliar with, please don’t hesitate to ask me to explain it. It will be interesting to see what analogies translate to other countries.

Start with a bad joke

Two statistics are sitting in a bar. One turns to the other and asks, “So, how do you like married life?”

The other statistic responds …

Put your reaction (“Ha ha”, “Groan”, etc.) in the chat box.

Speaker notes

One more thing efore I begin anything important. I like to start my talks with a silly joke. It always relates to something I am going to say later.

Now on Zoom, I often miss student reactions. So when I say something funny, I want you to type “Ha ha” or “Smile” or “LMFAO”. The acronym LMFAO means laughing my something … I forget how the rest of it goes.

Now if the joke is corny, like a really really bad pun, it’s okay to put “Groan”. The only thing bad is if I tell a joke and get no reaction at all.

I’ll be sneaking in some jokes throughout the talk and I really want a reaction from you, good or bad. If I don’t get any reaction to a bad pun, your “pun”ishment will be more bad puns.

So here’s the joke. It has been floating around on the Internet for quite a while, and I can’t find the person who gets credit for this. But here goes.

[Read joke and finish with] “It’s okay but you lose a degree of freedom.”

Okay, I’m waiting for reactions.

Introduction

- Tell us one interesting number about yourself
- Examples
  - 8: I have traveled to eight countries outside the United States
    - (Canada, Italy, China, France, Russia, England, Holland, and Iceland)
  - 29: I did not learn how to drive until I was 29 years old
  - 1802: My highest chess rating was 1802, but I am not that good any more.

Speaker notes

Speaker notes

I want to learn a bit about all of you, and I’m going to do this in a statistical way. Tell me one interesting number about yourself. It could be something simple, like the number of children you have or something exotic like the height of the highest mountain you have climbed.

Here are three numbers about me.

Your turn

A bit more about myself

- PhD in Statistics in 1982 from the University of Iowa
- Currently full professor
- Part-time statistical consultant
- Funded on 18 research grants
- Over 100 peer-reviewed publications
- Website with over 2,000 pages
- Many invitations to talk at conferences

Speaker notes

I have a PhD in Statistics from the University of Iowa. I have always had a strong interest in the computational side of Statistics. My dissertation was 150 pages, and 100 of those pages were computer generated graphs.

I am currently a full professor at the University of Missouri-Kansas City in the Department of Biomedical and Health Informatics. I also do statistical consulting on a part-time basis.

I have been a prolific researcher, receiving support from 18 different grants, and writing over 100 peer-reviewed publications.

I started a website in 1998, writing about data analysis, research ethics, and evidence based medicine. I wrote about two or three pages every week and my site now has over 2,000 pages. It shows the value of persistence.

I love to talk about Statistics and have given many presentations at regional, national, and international conferences. This ranges from short 15 minute talks to day long short courses.

Outline of the three day course

- Day one: Numerical summaries and data visualization
- Day two: Hypothesis testing and sampling
- Day three: Statistical tests to compare treatment to a control and regression models

My goal: help you to become a better consumer of statistics

Speaker notes

Speaker notes

This course will cover three days. Here are the topics for each day

Day one topics

- Numerical summaries
  - When should you present the mean versus the median
  - When should you present the range versus standard deviation
  - How should you display percentages
  - Why should you round liberally

Speaker notes

Speaker notes

Today, you will learn about numerical summaries.

Day one topics (continued)

- Data visualization
  - How should you display continuous data
  - Why is the normal bell-shaped curve important
  - How should you display categorical data
  - How do you illustrate trends and patterns
  - What are some common mistakes in the choice of colors

Speaker notes

Speaker notes

You will also learn a little bit about data visualization.

Quiz questions

Counting and proportions

- Counts are the most common statistic
  - Counts are error prone
  - Counts require a solid operational definition

Speaker notes

Speaker notes

Let’s start with the simplest statistic of all a simple count. This is probably the most common statistic produced.

But counts can be tricky. The counting process is error prone and requires a solid operational definition.

Student exercise

Count the number of occurrences of the letter “e”.

```
A quality control  program is easiest
to implement from the top down.
Make sure that you understand the
the commitment of time and money
that is involved. Every workplace is
different, but think about allocating
10% of your time and 10% of the
time of all your employees to
quality control.
```

Speaker notes

Speaker notes

Here’s an exercise I want you to do. Just count the number of occurrences of the letter “e”. Once you have your answer, type it in the chat box.

PAUSE HERE.

The numbers are different because of two things. First, it is easy to make mistakes. Did anyone notice the repetition of the word “the” at the end of the third line and the beginning of the fourth. It would be easy to miss that and count one less “e”.

What did you do with the first e in “Every”?

Did you count the e’s in the quotes itself or also on the slide instructions and the slide header?

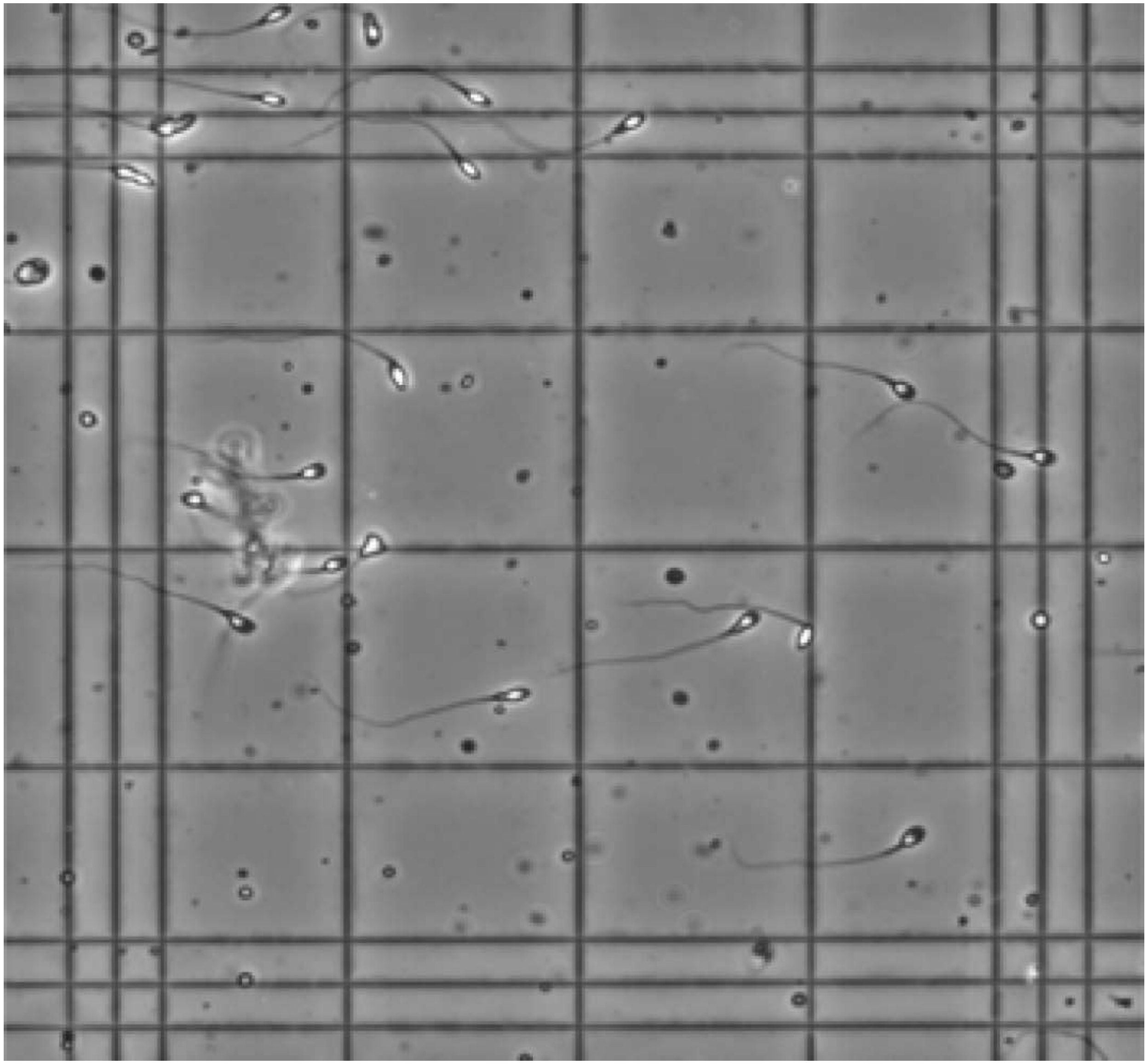


Figure 1: Image of a haemocytometer

Speaker notes

Speaker notes

This image is take from the WHO laboratory manual for the examination and processing of human semen, published in 2021. It shows a haemocytometer, an instrument used for counting the number of cells. To get a proper count, you need to include any cells inside the four by four grid of large squares in the middle of this micrograph. But what does “inside” mean? Should you count only those cells entirely inside the four by four grid. Or should you include cells that are partially inside the grid?

One rule is to count cells if the head of the sperm cell touches the top or right side of a square, but not if it touches the bottom or left side of the square. And don’t count a sperm cell if only the tail is inside the square.

That’s not the only way you can do this, but just make sure that whatever convention you use for deciding “inside” versus “outside” is consistent across your laboratory.

Sex * Survived Crosstabulation

Count

		Survived		
		No	Yes	Total
Sex	female	154	308	462
	male	709	142	851
Total		863	450	1313

Figure 2: Titanic data: counts of survival by gender

Speaker notes

Speaker notes

Here is some count data from an interesting data set. It shows who survived and who did not on the passenger ship, Titanic.

The Titanic was an enormous ship. It was bigger than any passenger ship ever built at the time. It was so large that they thought it was unsinkable. But in its first voyage across the Atlantic Ocean, it struck an iceberg and sunk.

They kept records on everyone on the ship: sex, age, and passenger class. There were 462 women on the ship. 308 of them survived, including Kate Winslet. The men did not fare as well. This was in a time when they really believed in the saying “Women and children first”. If this happened today, I’d push past all the ladies and the little kids and jump in that life boat first.

Among the 851 men, 709 died, including, sadly, Leonardo Di Caprio.

I’m making a reference to a popular movie, “Titanic” that was released in 1997. Has anyone seen that movie?

Anyway, you might want to examine mortality trends more closely by computing percentages. But there are three different ways you could compute these percentages.

Sex * Survived Crosstabulation

			Survived		
			No	Yes	Total
Sex	female	Count	154	308	462
		% within Survived	17.8%	68.4%	35.2%
	male	Count	709	142	851
		% within Survived	82.2%	31.6%	64.8%
Total		Count	863	450	1313
		% within Survived	100.0%	100.0%	100.0%

Figure 3: Titanic data with column percentages

Speaker notes

Speaker notes

Here are the percentages computed by dividing by the column totals. Divide the 308 surviving females by the total number of survivors, 450, to get 68%. Divide the 142 surviving males by 450 to get 32%. So those lifeboats were mostly, but not entirely, filled with women.

These are called column percents. They add up to 100% within each column: 18% + 82% = 100% and 68% + 32% = 100%.

Sex * Survived Crosstabulation

			Survived		
			No	Yes	Total
Sex	female	Count	154	308	462
		% within Sex	33.3%	66.7%	100.0%
	male	Count	709	142	851
		% within Sex	83.3%	16.7%	100.0%
Total		Count	863	450	1313
		% within Sex	65.7%	34.3%	100.0%

Figure 4: Titanic data with row percentages

Speaker notes

Speaker notes

You could also divide by the row totals. Divide the 308 surviving women by the total number of women, 462, to get a survival rate of 67%. Divide the 142 surviving men by the total number of men, 851, to get 17%.

!7%! This shows how poorly the men fared on the Titanic. If you were female, you might have died, but more likely than not you did survive. For the men, not such good news. Most of them died. Only a small fraction survived.

This is called the row percentages. These percentages add up to 100 within each row: 33% + 67% = 100% and 83% + 17% = 100%.

Percentages divided by grand total

Sex * Survived Crosstabulation

			Survived		
			No	Yes	Total
Sex	female	Count	154	308	462
		% of Total	11.7%	23.5%	35.2%
	male	Count	709	142	851
		% of Total	54.0%	10.8%	64.8%
Total		Count	863	450	1313
		% of Total	65.7%	34.3%	100.0%

Figure 5: Titanic data with cell percentages

Speaker notes

Speaker notes

You could also divide all the numbers by the grand total of 1,313. The 308 female survivors represented a bit less than 24% of all the passengers that set sail from England.

The 142 male survivors represented a bit less than 11% of all the survivors.

These are called the cell percentages. They add up to 100% across the entire table: 12% + 54% + 24% + 11% = 101%. Close enough!

Which makes the most sense? It depends on your perspective. If you want to test the hypothesis that male passengers on the Titanic had a much smaller risk of dying, then the row percentages make the most sense.

But from the perspective of the Carpathia, the ship that rescued the survivors, the column percents make the most sense. They had to make room on their ship for 450 passengers, 68% who were female and 32% who were male. I bet that the lines for the women’s bathrooms on the Carpathia were really long.

My recommendations

- Treatment or exposure as rows
- Outcome as columns
- Usually report row percentages
  - Female survival rate: 67%
  - Male survival rate: 17%
- But sometimes column percentages
  - Survivors: 68% female, 32% male

Speaker notes

Speaker notes

I have some general guidelines that I use. They don’t always work, but they work most of the time.

If you have a variable that represents a treatment or exposure, try using that as the rows of the table. If you have a variable that represents an outcome, try using that as the columns of the table. Sometimes, there are no clearly identified treatment variables and no clearly identified outcome variables. But try to categorize them this way, if you can.

With a table lined up with the treatments as the rows and the outcomes are the variables, calculate the row percentages.

In the Titanic data, survival is clearly an outcome. So arrange the table like I did with sex as the rows and survival as the columns and compare the two survival rates: a healthy 67% for females and a feeble 17% for males.

But sometimes you will find that the column percents make more sense. It does depend on what question you are trying to answer with the data.

Some rationale for these choices

My way

```
             Survived
             No          Yes
Sex Female   33% (154)   67% (308)
    Male     83% (863)   17% (142)
```

Not my way

```
                  Sex
                  Female      Male
Survived   No     33% (154)   83% (863)
           Yes    67% (308)   17% (142)
```

Speaker notes

Speaker notes

Now, I believe it is important to think carefully about which is your rows and which is your columns. Here’s the layout that I recommend on the left and the layout that I don’t recommend on the right. The key comparison is among survival rates, 67% for females and only 17% for males. When you orient my way with the treatment/exposure (Sex) as rows and the outcome (Survived) as the columns, the numbers 67% and 17% are very close to one another. In the alternate layout the numbers you are most interested in comparing are not as close together.

Now this is not an absolute rule. Sometimes I’ll switch things up. But about 90% of the time, I find that the layout with the treatment or exposure as the rows and the outcome as the columns, the table just looks better.

Break

- What have you just learned?
    - Displaying percentages
- What is coming next?
    - Practice exercise
    - Calculation of the mean and median

Speaker notes

Speaker notes

Let me pause here for a second. Are there any questions?

On your own

Calculate row and column percentages for the following tables. Interpret your results.

PClass * Survived Crosstabulation

Count

		Survived No	Survived Yes	Total
PClass	1st	129	193	322
	2nd	161	119	280
	3rd	573	138	711
Total		863	450	1313

Figure 6: Titanic passenger class counts

Child * Survived Crosstabulation

Count

		Survived No	Survived Yes	Total
Child	No	386	244	630
	Yes	57	69	126
Total		443	313	756

Figure 7: Titanic child counts

Speaker notes

Speaker notes

Now try to report both column and row percents for one of these two tables. Breakout room #1 work on the passenger class table and breakout room #2 work on the child data.

Put your percentages in a table using a word processing program or text editor so you can share your results with the group.

Be sure to interpret these numbers. Come back together again in about 10 minutes.

Figure 8: Cartoon image of Professor Mean

Speaker notes

Speaker notes

Here’s a cartoon image of Professor Mean. I know this looks like it was drawn by a professional artist, but it was actually drawn by me. Really!

Professor Mean is my alter ego on the Internet. For those who don’t get the inside joke, I point out that Professor Mean is not just your average professor.

I will use the terms mean and average interchangeably througout this talk.

Figure 9: Road with a median strip

Speaker notes

Speaker notes

This is an image of a traffic median. This is a strip of land, typically raised from the road surface, that splits the road in half.

In Statistics, the median is the data value that splits the data in half. Half of the data is smaller than the median and half of the data is larger than the median.

Calculation of the mean and median

- Mean
  - Add up all the values, divide by the sample size
- Median
  - Sort the data
    - Select the middle value if n is odd
    - go halfway between the two middle values if n is even

Speaker notes

Speaker notes

You already know how to compute the average. Add up all the values and divide by the sample size.

The median is also simple. Sort the data and choose the “middle” value. If n is odd, there is one value that is right in the middle. With five data values, the median is the third value of the sorted list. The first and second values are smaller and the fourth and fifth values are larger.

With an even number, there are two middle values. Go halfway between them. If you have eight data values, the midpoint between the fourth and fifth values splits the data in half. The first through fourth values in the sorted list are smaller and the fifth through eighth values are larger.

Formal mathematical definitions

- Mean
  - $\bar{X} = \frac{1}{n} \Sigma X_i$
- Median
  - Sorted values $X_{[1]}, X_{[2]}, \ldots, X_{[n]}$
    - $X_{[(n+1)/2]}$ if n is odd,
    - $(X_{[n/2]} + X_{[n/2+1]})/2$ if n is even

Speaker notes

Speaker notes

Here are the mathematical formulas for the mean and median. I know some people hate formulas, but I love them. With a few symbols and Greek letters, you can express really deep and beautiful ideas. Well these formulas aren’t all that deep.

Bacteria before and after A/C upgrade

Room	Before	After	Change
121	11.8	10.1	-1.7
125	7.1	3.8	-3.3
163	8.2	7.2	-1.0
218	10.1	10.5	0.4
233	10.8	8.3	-2.5
264	14	12	-2.0
324	14.6	12.1	-2.5
325	14	13.7	-0.3

Speaker notes

Speaker notes

Before remediation mean

```
11.8 + 7.1 + 8.2 + 10.1 + 10.8 + 14 + 14.6 + 14 = 90.6

90.6 / 8 = 11.325

Round to 11.3
```

Speaker notes

Speaker notes

Here's the data for bacterial counts before remediation. If you add the eight values up, you get 90.6. Divide this by eight to get 11.325. Always round liberally when you are talking about the mean.

After remediation mean

```
10.1 + 3.8 + 7.2 + 10.5 + 8.3 + 12 + 12.1 + 13.7 = 77.7

77.7 / 8 = 9.7125

Round to 9.7
```

Speaker notes

Speaker notes

Here are the same calculations for the bacterial counts after remediation.

Before remediation median (1/4)

121	11.8
125	7.1
163	8.2
218	10.1
233	10.8
264	14.0
324	14.6
325	14.0

Speaker notes

Speaker notes

Here is the data for bacteria counts before remediation. Notice that the data is arranged by room number.

Before remediation median (2/4)

```
125   7.1
163   8.2
218  10.1
233  10.8
121  11.8
264  14.0
325  14.0
324  14.6
```

Speaker notes

Speaker notes

The first thing you do is sort the data from the lowest bacteria count to the highest bacteria count.

The data was arranged by toom number, but now it is arranged by bacterial count.

Before remediation median (3/4)

```
125   7.1
163   8.2
218  10.1
233  10.8  10.8
121  11.8  11.8
264  14.0
325  14.0
324  14.6
```

Speaker notes

Speaker notes

Then pick out the middle value. If you have an even number of data points, there will be two middle values.

In this data set, the two middle values are the fourth and fifth largest values out of eight.

Before remediation median (4/4)

```
125   7.1

163   8.2

218  10.1

233  10.8  10.8
                  (10.8 + 11.8) / 2 = 11.3
121  11.8  11.8

264  14.0

325  14.0

324  14.6
```

Speaker notes

Speaker notes

If there are two middle values, just average them.

After remediation median (1/4)

121	10.1
125	3.8
163	7.2
218	10.5
233	8.3
264	12.0
324	12.1
325	13.7

Speaker notes

Speaker notes

Here is the data for bacteria counts after remediation.

After remediation median (2/4)

```
125   3.8
163   7.2
233   8.3
121  10.1
218  10.5
264  12.0
324  12.1
325  13.7
```

Speaker notes

Speaker notes

Just like before, you sort the data.

After remediation median (3/4)

```
125   3.8
163   7.2
233   8.3
121  10.1  10.1
218  10.5  10.5
264  12.0
324  12.1
325  13.7
```

Speaker notes

Speaker notes

Then pick out the middle value. Here again, there are two middle values.

After remediation median (4/4)

```
125   3.8

163   7.2

233   8.3

121  10.1  10.1
                    (10.1 + 10.5) / 2 = 10.3
218  10.5  10.5

264  12.0

324  12.1

325  13.7
```

Speaker notes

Speaker notes

Just average the two middle values.

Break

- What have you just learned?
  - Calculation of the mean and median
- What is coming next?
  - Criticisms of the mean and median

Speaker notes

Speaker notes

Let me pause here for a second. Are there any questions?

Criticisms of the mean and median

- Are you combining apples and onions?
- Are you ignoring minorities?

Speaker notes

There’s a wonderful cartoon by Dana Fradon that appeared in The New Yorker in 1976. She shows a road going into town and the sign by the side of the road reads “Hillsdale, Founded 1802, Altitude 600, Population 3,700. Total 6,122.” You can’t add these things together.

It’s similar for means. There was a dataset showing housing prices for homes in Boston and none of the analyses seemed to make sense. The problem in Boston is that a small number of the houses had prices that were out of sync with their other homes. These were historical houses, such as Paul Revere’s house.

When you are averaging numbers, maybe it’s okay to have a few oranges in with the apples. A mix of apples and oranges is just fruit salad. You shouldn’t have a problem with that.

When it becomes a problem is when the data are so diverse that it becomes a mix of apples and onions. There are lots of great recipes that mix apples and oranges, but none that mix apples and onions.

The other problem is that an average may be a reasonable number to represent the majority of patients in your sample, but it may masks some important trends that appear in a minority.

This is a big problem in a larger context than just the mean or median. There are some very fancy high tech predition models that work very well for most people and the statistics like the mean and median back this up quite nicely. But the prediction models perform terribly for minority groups.

Use of the mean for ordinal data

- Stevens scales of measurement (controversial!)
  - Nominal
  - Ordinal
  - Interval
  - Ratio
- Addition/subtraction not allowed for ordinal data
  - Mean of ordinal data is meaningless

Speaker notes

Speaker notes

A psychologist, Stanley Smith Stevens divided the entire universe of data into four categories: nominal, ordinal, interval, ratio. I won’t review the definitions for all of these, but ordinal data is categorical data where there is a natural ordering of categories. An important limitation to ordinal data, but where the spacing between successive units is not consistent.

The belief among many (but not all) researchers, is that

An example of ordinal data.

- “Do you agree or disagree with the following statements”
  - “I believe that knowledge of Statistics is important for my job.”
    - 1 = Strongly disagree,
    - 2 = Disagree
    - 3 = Neutral
    - 4 = Agree
    - 5 = Strongly agree

Speaker notes

Speaker notes

An example of ordinal data is the Likert scale. This takes various forms, but often it is used with group of questions on a questionnaire that reads something like

“Do you agree or disagree with the following statements”

You are asked to respond 1=Strongly disagree, 2=Disagree, 3=Neutral, 4=Agree, 5=Strongly agree.

Now I’m sure everyone today is going to choose 5. But assigning numbers 1, 2, 3, 4, and 5 to categories of strongly disagree, disagree, neutral, agree, and strongly agree may falsely imply that a jump from 3 (neutral) to 4 (agree) is about the same amount of improvement as a jump from 4 (agree) to 5 (strongly agree). That’s probably not the case.

You can’t really average ordinal data, some people say because that implies that two responses of “Agree” are the same as one response of “Neutral” along with a response of “Strongly agree”.

Do you want everyone to be at least somewhat on your side or do you want to have a smaller number of very enthusiastic supporters.

If you believe that two 4’s are not the same as a 3 and a 5, then you can’t average.

Now I beg to disagree here, but I am part of a minority opinion. I think that if at the start of this class, your average rating was 3.2 and after I finish the lecture, your average rating climbs to 4.4, that I have done my job well.

If it only jumps to 3.6, then I have still done well, but not as much as that jump to 4.4.

Another example of ordinal data, course grades

- A = 4
- B = 3
- C = 2
- D = 1
- F = 0

Speaker notes

Speaker notes

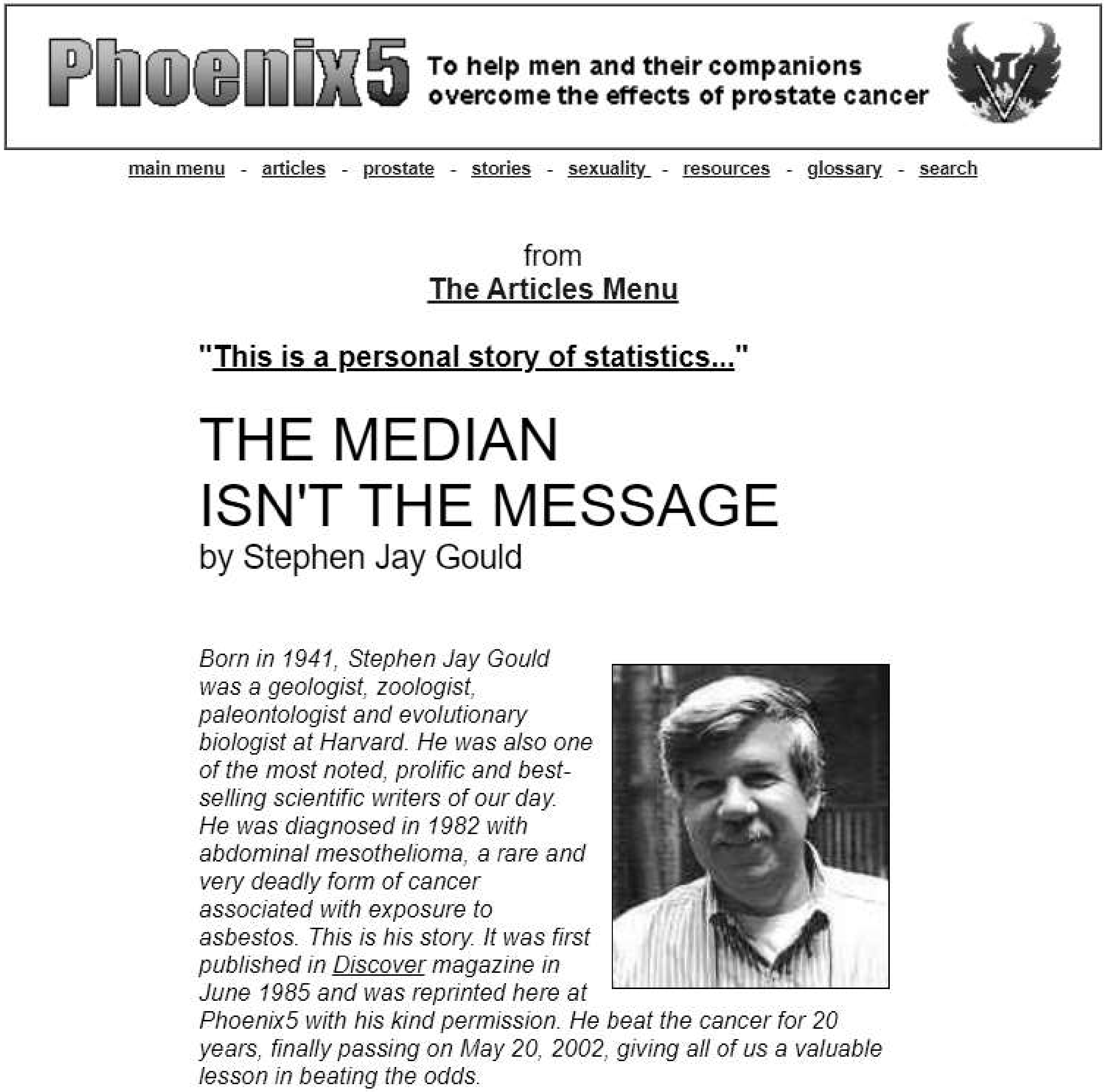
Another example of ordinal data is grades assigned to students. Now everyone in this class is getting an A, but in other classes I teach I might assign different grades. You can attach a number to each of these grades, 4 for A, 3 for B, 2 for C, 1 for D, and 0 for F.

These numbers seem to imply that a student with two B's is as smart as a student with an A and a C.

It raises an interesting story. A colleague of mine told me that he would never hire anyone with a single F on their transcript. An F is a red flag, he felt. So he would not want to assign a value of 0 to F, because that implies that the difference between an F and a D is equivalent to the difference between a B and and A. He's want to assign a value like negative one million to an F so that the average would be pulled way down for a single F, no matter what the other grades would be.

Now I would never be so harsh, but there is really nothing wrong with his perspective. And I would certainly treat a student with three A's and one F differently from a student with two A's and two C's even though mathematically, both average out to 3.0.

Now, in spite of all the obvious problems with equivalence between different grades, most of us still accept a grade point average as a meaningful indicator of how well a student did in school.



Phoenix5 To help men and their companions overcome the effects of prostate cancer

main menu - articles - prostate - stories - sexuality - resources - glossary - search

from
The Articles Menu

"This is a personal story of statistics..."

THE MEDIAN
ISN'T THE MESSAGE
by Stephen Jay Gould

Born in 1941, Stephen Jay Gould was a geologist, zoologist, paleontologist and evolutionary biologist at Harvard. He was also one of the most noted, prolific and best-selling scientific writers of our day. He was diagnosed in 1982 with abdominal mesothelioma, a rare and very deadly form of cancer associated with exposure to asbestos. This is his story. It was first published in Discover magazine in June 1985 and was reprinted here at Phoenix5 with his kind permission. He beat the cancer for 20 years, finally passing on May 20, 2002, giving all of us a valuable lesson in beating the odds.

Figure 10: Excerpt from Gould 1985 publication

Speaker notes

Speaker notes

Stephen Jay Gould was a famous Evolutionary Biologist. He was a prolific writer with 20 books and 300 essays. Much of his writing was for academic researchers, but just as much was for the general public.

One of his most famous essays was “The Median Isn’t the Message”. The title is a take-off of a quote by Marshall McLuhan, “The medium is the message” which itself has an interesting history that you should investigate on your own.

The Gould essay was written in 1985 for Discover Magazine. It has been reprinted many times, and you can easily find the full text with a simple Google search.

The image shown here is taken from phoenix5.org, an informational site for patients with prostate cancer.

Choosing between the mean and median

- Often a source of controversy
- When do you use the mean?
  - When totals are important
- When do you use the median
  - When outliers/skewness might distort your conclusions
- Often, either is fine

notes

Speaker notes

While there is some consensus on when to use the mean versus the median, the choice is not always obvious. Controversies often arise over this issue.

Here are some general guidelines.

The mean allows extrapolation to totals. This is often important in the analysis of the economic effects of illness. :::

Eur Respir J 2001; 17: 495–498

European Respiratory Journal
ISSN 0903-1936

Airway resistance measured by the interrupter technique: expiration or inspiration, mean or median?

P.D. Bridge, S.A. McKenzie

Figure 11: Exceprt from Bridge and McKenzie 2001, PMID: 11405531

44

Speaker notes

Speaker notes

Bridge 2001, PMID: 11405531 (continued)

The measurement of airway resistance by the interrupter technique (Rint) needs standardization. Should measurements be made be during the expiratory or inspiratory phase of tidal breathing? **In reported studies, the measurement of Rint has been calculated as the median or mean of a small number of values, is there an important difference?**

45

Speaker notes

Speaker notes

Bridge 2001, PMID: 11405531 (continued)

In the present data the mean of a set of values contributing to a measurement was not significantly different from the median. **However, the use of the median has been recommended since it is less affected by possible outlying values such as might be included by fully automated equipment.**

Speaker notes

Speaker notes

Chen 2019, PMID: 31806195

Author Manuscript

HHS Public Access

Author manuscript

Value Health. Author manuscript; available in PMC 2020 December 01.

Published in final edited form as:

Value Health. 2019 December ; 22(12): 1387–1395. doi:10.1016/j.jval.2019.08.005.

Trends in the Price per Median and Mean Life-Year Gained Among Newly Approved Cancer Therapies 1995 to 2017

Alice J. Chen, PhD[1,2,*], Xiaohan Hu, MPH[2], Rena M. Conti, PhD[3], Anupam B. Jena, MD, PhD[4,5], Dana P. Goldman, PhD[1,2,5]

Figure 12: Chen et al 2019

Speaker notes

Speaker notes

Chen 2019, PMID: 31806195 (continued)

Background: The prices of newly approved cancer drugs have risen over the past decades. **A key policy question is whether the clinical gains offered by these drugs in treating specific cancer indications justify the price increases.**

Speaker notes

Speaker notes

Chen 2019, PMID: 31806195 (continued)

Results: We found that between 1995 and 2012, price increases outstripped median survival gains, a finding consistent with previous literature. **Nevertheless, price per mean life-year gained increased at a considerably slower rate, suggesting that new drugs have been more effective in achieving longer-term survival.** Between 2013 and 2017, price increases reflected equally large gains in median and mean survival, resulting in a flat profile for benefit-adjusted launch prices in recent years.

Speaker notes

Speaker notes

Percentiles

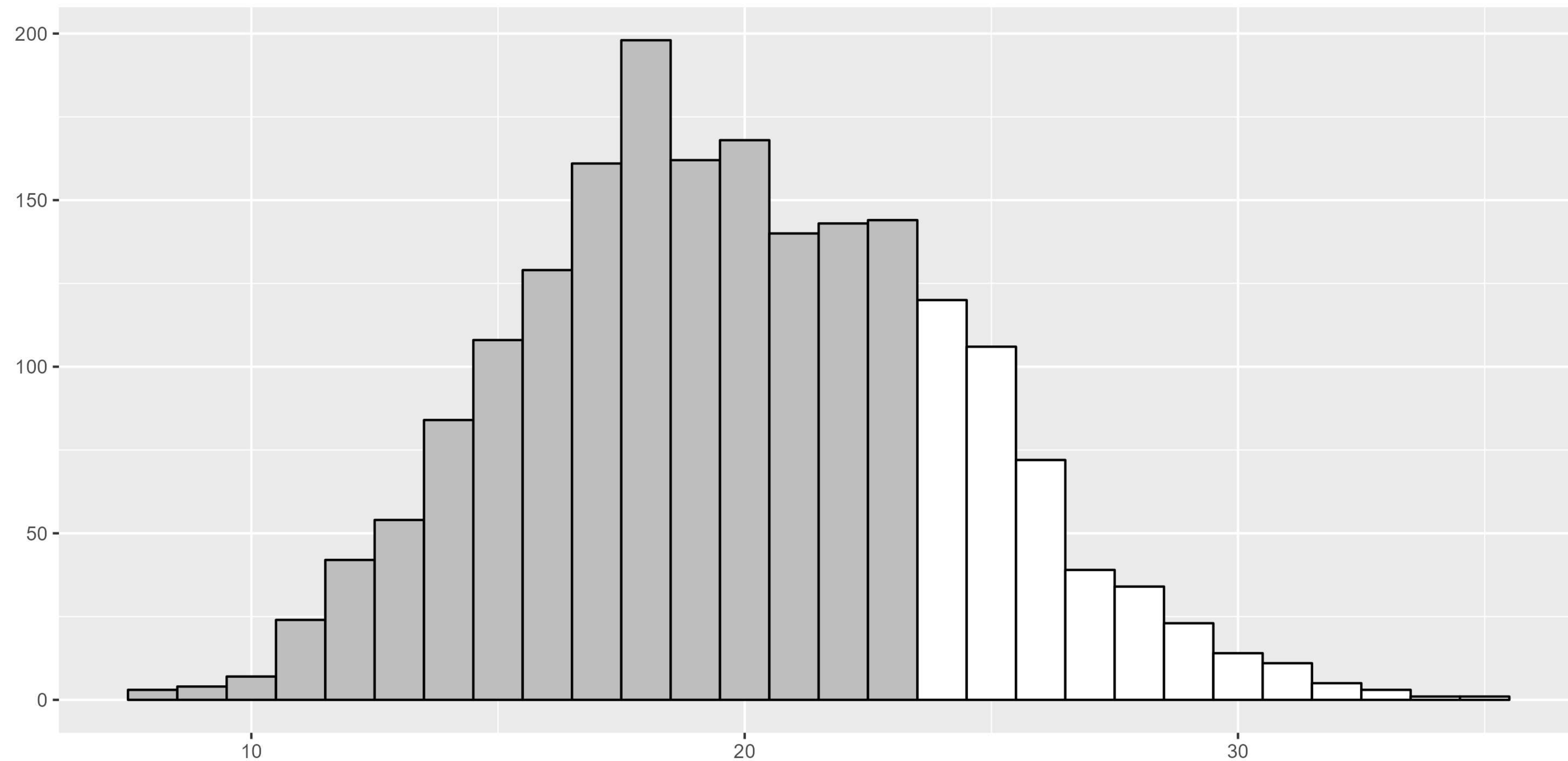


Figure 13: Illustration of the 75th percentile

Speaker notes

Speaker notes

I want to mention percentiles briefly. A percentile is a value that splits the data so that a certain percentage is smaller and a certain percentage is larger.

The 75th percentile, for example will be above 75% of the data and below 25% of the data. This graph illustrates the 75th percentile for some arbitrary data. THe gray bars represent about 75% of the data and the white bars represent about 25% of the data.

I use a few weasel words like “roughly” and “about” because you can’t always get a perfect split. But you can usually come close.

Computing percentiles

- Many formulas
  - Differences are not worth fighting over
- My preference (pth quantile)
  - Sort the data
  - Calculate p*(n+1)
  - Is it a whole number?
    - Yes: Select that value, otherwise
    - No: Go halfway between
    - Special cases: p(n+1) < 1 or > n

Speaker notes

Speaker notes

There are close to a dozen different ways to compute a percentile, but the differences between the values selected are small and not worth fussing about.

Here is my preference for choosing the pth quantile (remember that for quantiles, you range between 0 and 1, not between 0 and 100).

Calculate the quantity $p*(n+1)$. If that value is a whole number, great! You just select that value. If it is a fractional value, round up and down and go halfway between.

Once in a while, you'll get an extreme case, where $p(n+1)$ is less than 1 or greater than n. Just use a bit of common sense.

If you have nine values and $p(n+1)$ is 9.2, you can't go halfway between the 9th and 10th observations. There is no 10th observation. So just choose the 9th or largest value.

Likewise if $p(n+1)$ is 0.8, you can't go halfway between the zeroth and first observation. There is no zeroth observation. Just choose the first or smallest value.

Some examples of percentile calculations

- Example for $n=39$
  - For 5th percentile, $p(n+1)=2$ -> 2nd smallest value
  - For 4th percentile, $p(n+1)=1.6$ -> halfway between two smallest values
  - For 2nd percentile, $p(n+1)=0.8$ -> smallest value

Speaker notes

Speaker notes

Suppose you have 39 observations. For the 5th percentile or the 0.05 quantile, $p(n+1)$ equals 2. Lucky you. The second smallest observation is the 5th percentile. For the 4th percentile or the 0.04 quantile, you get $p(n+1)$ equal to 1.6. Go halfway between 1, the smallest value, and 2, the second smallest value.

The 2nd percentile represents one of the special cases. You calculate $p(n+1)$ and get 0.8. You can't go halfway between 0 and 1, so just choose the smallest value.

Some terminology

- Percentile: goes from 0% to 100%
- Quantile: goes from 0.0 to 1.0
  - 90th percentile = 0.9 quantile
- Quartiles: 25th, 50th, and 75th percentiles
  - Lower quartile: 25th percentile
  - Upper quartile: 75th percentile

Speaker notes

Speaker notes

A percentile always refers to a percentage. So it has to be between 0% and 100%. Sometimes, you may see references to a quantile. A quantile is a percentile, but is expressed as a proportion rather than a percent. A quantile goes from 0.0 to 1.0. The 25th percentile and the 0.25 quantile are the same thing.

You might see the term “quartiles”. These are the 25th, 50th, and 75th percentiles. These three values split the data into quarters.

If you see “lower quartile”, it means the 25th percentile. Likewise, “upper quartile” means the 75th percentile.

Let me be try to be careful about terminology here. But, sometimes I will mess up and use “percentile” when I mean “quantile”.

Before remediation upper quartile (1/4)

```
121  11.8
125   7.1
163   8.2
218  10.1
233  10.8
264  14.0
324  14.6
325  14.0
```

Speaker notes

Speaker notes

Here is the data for bacteria counts before remediation. Let’s calculate the upper quartile, also known as the 0.75 quantile or the 75th percentile.

Before remediation upper quartile (2/4)

```
125   7.1
163   8.2
218  10.1
233  10.8
121  11.8
264  14.0
325  14.0
324  14.6
```

Speaker notes

Speaker notes

Just like before, you sort the data.

Before remediation upper quartile (3/4)

```
125   7.1

163   8.2

218  10.1

233  10.8

121  11.8

264  14.0  14

325  14.0  14

324  14.6
```

Speaker notes

Speaker notes

With $n=8$, you get $p(n+1) = 6.75$. So pick out the sixth and seventh values.

Before remediation upper quartile (4/4)

```
125   7.1

163   8.2

218  10.1

233  10.8

121  11.8

264  14.0  14
                  (14 + 14) / 2 = 14
325  14.0  14

324  14.6
```

57

Speaker notes

Speaker notes

Go halfway between these values.

After remediation upper quartile (1/4)

```
121  10.1

125   3.8

163   7.2

218  10.5

233   8.3

264  12.0

324  12.1

325  13.7
```

Speaker notes

Speaker notes

Here are the same calculations the upper quartile of bacteria counts after remediation.

After remediation upper quartile (2/4)

```
125   3.8
163   7.2
233   8.3
121  10.1
218  10.5
264  12.0
324  12.1
325  13.7
```

Speaker notes

Speaker notes

Just like before, you sort the data.

After remediation upper quartile (3/4)

```
125   3.8
163   7.2
233   8.3
121  10.1
218  10.5
264  12.0  12
324  12.1  12.1
325  13.7
```

Speaker notes

Speaker notes

Then pick out the sixth and seventh values.

After remediation upper quartile (4/4)

```
125   3.8

163   7.2

233   8.3

121  10.1

218  10.5

264  12.0  12
                    (12 + 12.1) / 2 = 12.05
324  12.1  12.1

325  13.7
```

61

Speaker notes

Speaker notes

Go halfway between these values.

When you should use percentiles

- Characterize variation
  - Middle 50% of the data
- Exposure issues
  - Not enough to control median exposure level
- Quantify extremes
  - What does “upper class” mean?
- Quality control
  - Almost all products must meet a minimum standard

Speaker notes

Speaker notes

There are many reasons why you might be interested in percentiles rather than the mean or median. Actually, the median is a percentile, the 50th percentile, but what I mean is percentiles other than 50%.

One important use of percentiles is looking at the middle 50% of the data. This is the data between the lower quartile (25th percentile) and the upper quartile (75th percentile). Is the middle 50% of the data bunched tightly together or spread widely apart?

Percentiles are also important in the study of exposures. If you work in an environment where the median worker has a safe level of exposure, you could easily end up with 20%, 30% or more of the workers dying from unsafe exposures. It is important to insure that not just the median, but a very high percentile like the 99th percentile of exposure levels is at a safe level.

Percentiles also help to define extreme groups. You can, for example, define the term upper class as anyone earning more than the 90th percentile of income.

Percentiles also can help with quality control. If you make a claim about a product, you want to make sure that that claim is not valid at a median level but at a much higher level. You don't sell 500 mg bottles of liquid Tylenol is your factory is churning out a median fill level of 500 mg. Half of your customers would be cheated. Instead you insure that the 98th percentile coming out of the factory floor is at least 500 mg. You lose a bit of money because most bottles contain more than 500 mg, but the cost of an irate customer is worth more than the cost of 50 overfilled bottles.

Standard deviation

$$S = \sqrt{\frac{1}{n-1}\Sigma(X_i - \bar{X})^2}$$

At least one alternative formulas.

Speaker notes

Speaker notes

The standard deviation is a commonly used measure of how spread out the data is. The formula is a bit messy, but if you look carefully at it, you will see that it is a measure of how far each individual value is from the overall mean.

Now, maybe you've seen or used a different formula. Don't worry about it. In a short course like this, I won't ask you to calculate anything as tedious as a standard deviation. Let the computer do all of the work.

Why is variation important

- Variation = Noise
  - Too much noise can hide signals
- Variation = Heterogeneity
  - Too little heterogeneity, hard to generalize
  - Too much heterogeneity, mixing apples and oranges
- Variation = Unpredictability
  - Too much unpredictability, hard to prepare for the future
- Variation = Risk
  - Too much risk can create a financial burden

Speaker notes

Speaker notes

I want to discuss measures of variation now. Variation gets at the heart and soul of clinical statistics. A large portion of statistical analysis involves characterizing variation.

Variation can be thought of as a measure of noise. In general, but not always, noise is bad. Consider measuring a patient’s glucose level, to see if you have early evidence of diabetes. Your glucose level varies a lot during the day based on whether you skipped breakfast or decided to get a mid-afternoon Snickers bar. Your glucose level is noisy. A high level might or might not mean trouble. A low value might or might not mean you are safe. The large standard deviation of your measures of blood glucose indicates noise.

That’s why you are asked to take an overnight fast before testing your blood glucose level. Controlling your diet by not eating anything after midnight provides a more consistent measure of blood glucose. It has a smaller standard deviation and a high or low value is more helpful in diagnosis.

Variation can also be thought of as a measure of heterogeneity. Heterogeneity is also bad sometimes, but there are times when you want a fair amount of heterogeneity. A research study that has a lot of variation is better at providing a complete picture of what a typical patient is. Outcomes that are consistent in the presence of demographic heterogeneity give you more confidence in generalizing the results of a research study. You have some assurance that the therapy is not restricted to helping a small segment of patients.

Too much heterogeneity, though, can mean that any summary measure is a mixture of apples and oranges. You have to find the right balance.

Variation can be equated to unpredictability. The number of beds needed in a hospital does vary, and this makes it difficult to staff properly. The more variation in beds needed, the more headaches you have.

Variation can also be equated to risk. If you invest in a new drug, paying millions or even billions of dollars in testing, you are doing so with the hope that your investment will pay off. Unfortunately, the market for your drug is uncertain, and you might end up with no market at all if your clinical trials fail to convince FDA. There is variation in the return on your investment, and the more variation there is, the more risky your development plans are.

Should you try to minimize variation?

- Yes, for early studies
  - Easier to detect signals
  - Proof of concept trials
- No, for later studies
  - Easier to generalize results
  - Pragmatic trials

Speaker notes

Speaker notes

It is a bit of a generalization, but most researchers try to avoid variation in early studies. By early studies, I mean studies of therapies that have not yet been extensively tested in a broad range of settings. Less variation means that there is a greater chance to detect signals. You remove variation by using very strict entry criteria on who can get into the study. You remove variation by tightly controlling what the patient is allowed to do (e.g., no concommitant medications). You remove variation by tightly standardizing the delivery of the intervention and the assessment of the outcome. You reduce variation by removing patients who deviate from the research protocol requirements.

These are known as proof of concept trials. If a new therapy cannot succeed even under the tight controls, there is no point in studying it futher. But success in a tightly controlled environment does not guarantee success in the real world.

If you are planning a trial that comes after many similar trials, you actually may want to encourage variation. Broaden the inclusion criteria so that the patients in the trial look no different than the patients you see every day in your clinic.

Standard deviation

$$S = \sqrt{\frac{1}{n-1}\Sigma(X_i - \bar{X})^2}$$

At least one alternative formulas.

Speaker notes

Speaker notes

The standard deviation is a commonly used measure of how spread out the data is. The formula is a bit messy, but if you look carefully at it, you will see that it is a measure of how far each individual value is from the overall mean.

Now, maybe you've seen or used a different formula. Don't worry about it. In a short course like this, I won't ask you to calculate anything as tedious as a standard deviation. Let the computer do all of the work.

The bell shaped curve

- Does your variation follow a bell shaped curve?
  - Values in the middle are most common
  - Frequencies taper off away from the center
  - Symmetry on either side
- A bell shaped curve = better characterization of variation

Speaker notes

Speaker notes

Much variation in the real world follows a bell shaped curve, alternately called a normal distribution. You can assess whether you have a bell shaped curve using a histogram. Look for values in the middle being most common. The frequencies should taper off slowly as you moved away from the middle. The histogram should have symmetry. The left side of the histogram should be roughly equivalent to the right side of the histogram.

Not a bell shaped curve (1/4)

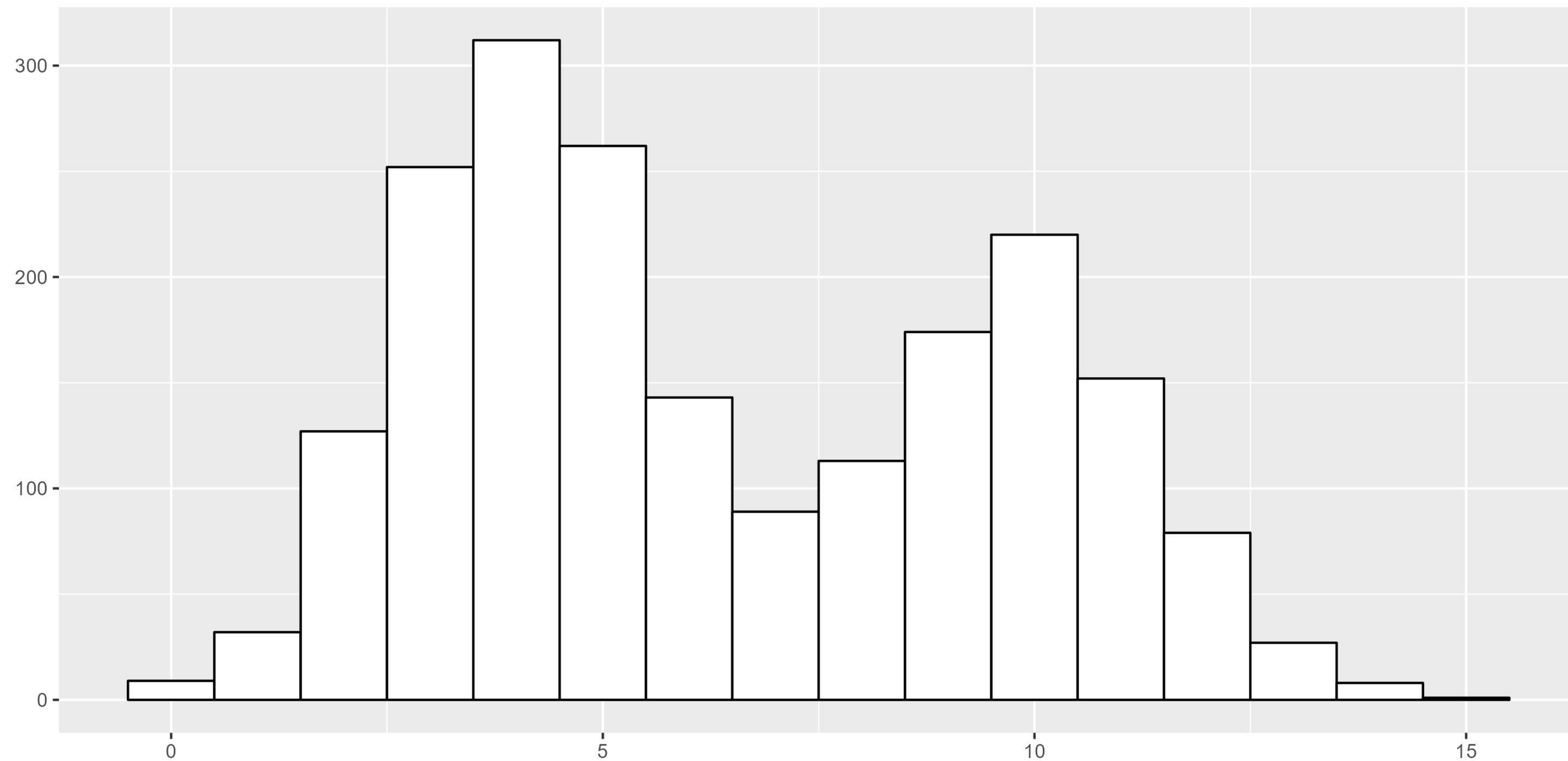


Figure 14: Bimodal histogram

68

Speaker notes

Speaker notes

Here’s a histogram that shows a bimodal distribution. The frequencies are not highest in the center of the data. This is not a bell shaped curve.

Not a bell shaped curve (2/4)

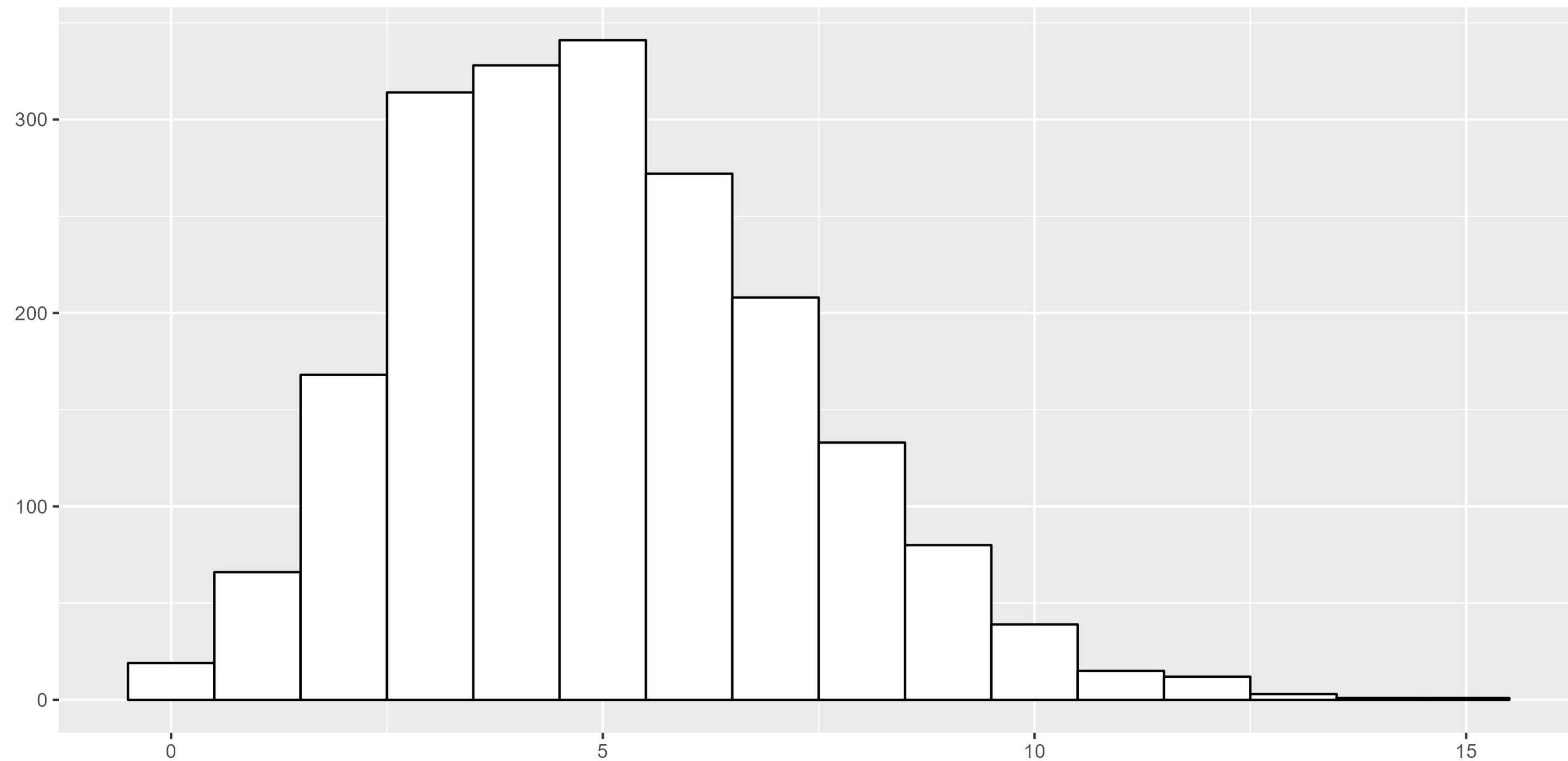


Figure 15: Skewed histogram

69

Speaker notes

Speaker notes

Here’s a histogram that shows a skewed or asymmetric distribution. This is not a bell shaped curve.

Not a bell shaped curve (3/4)

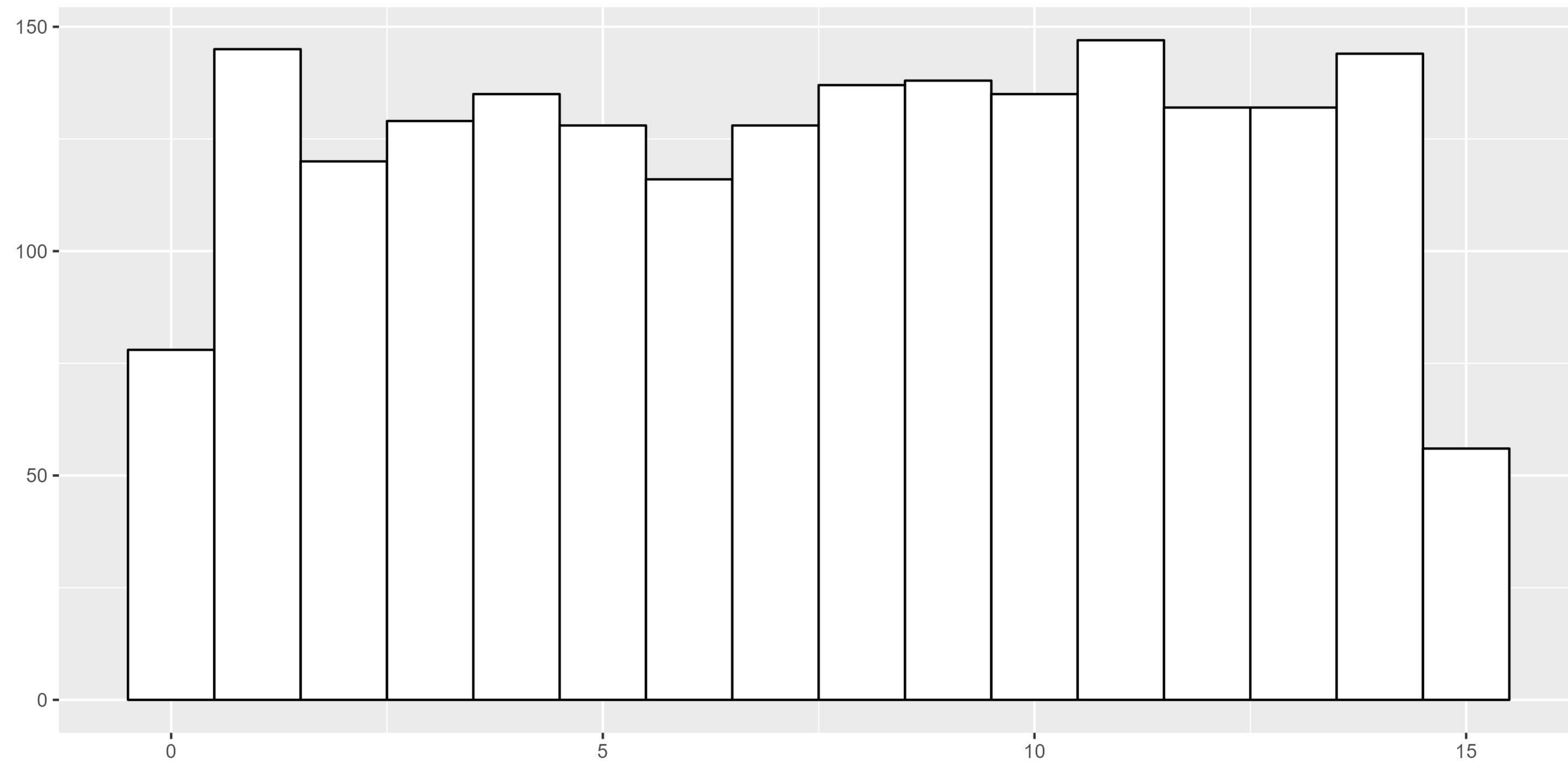


Figure 16: Uniform histogram

Speaker notes

Speaker notes

Here’s a histogram that shows a symmetric distibution, but the frequencies do not taper off as you move away from the center. This is not a bell shaped curve.

Not a bell shaped curve (4/4)

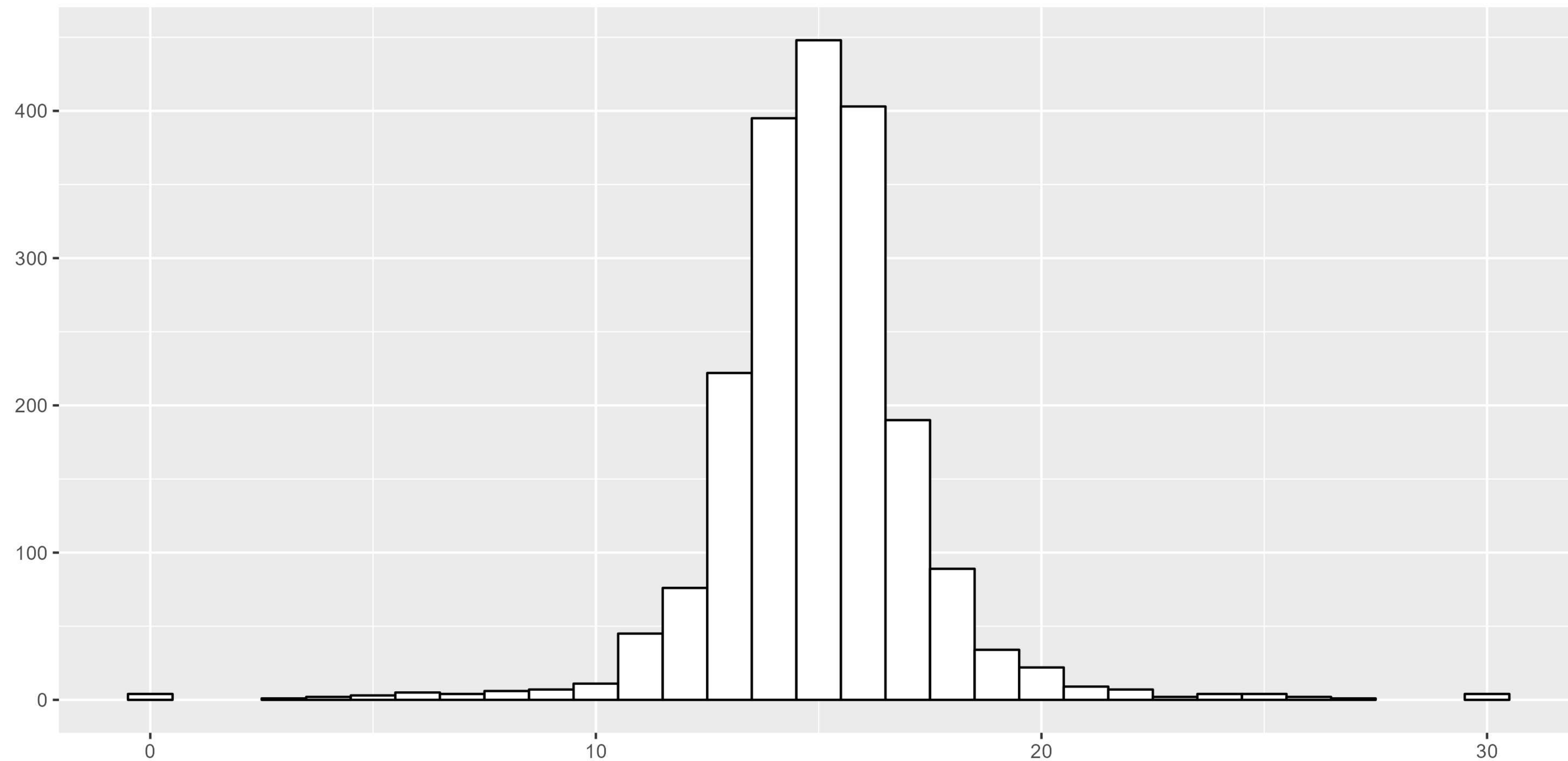


Figure 17: Heavy-tailed histogram

71

Speaker notes

Speaker notes

Here’s a histogram that shows a symmetric distibution, but the frequencies taper off at first, but then flatten out. This is called a heavy tailed distribution and it tends to produce outliers, extreme values, on both sides. This is not a bell shaped curve.

A bell shaped curve (finally!)

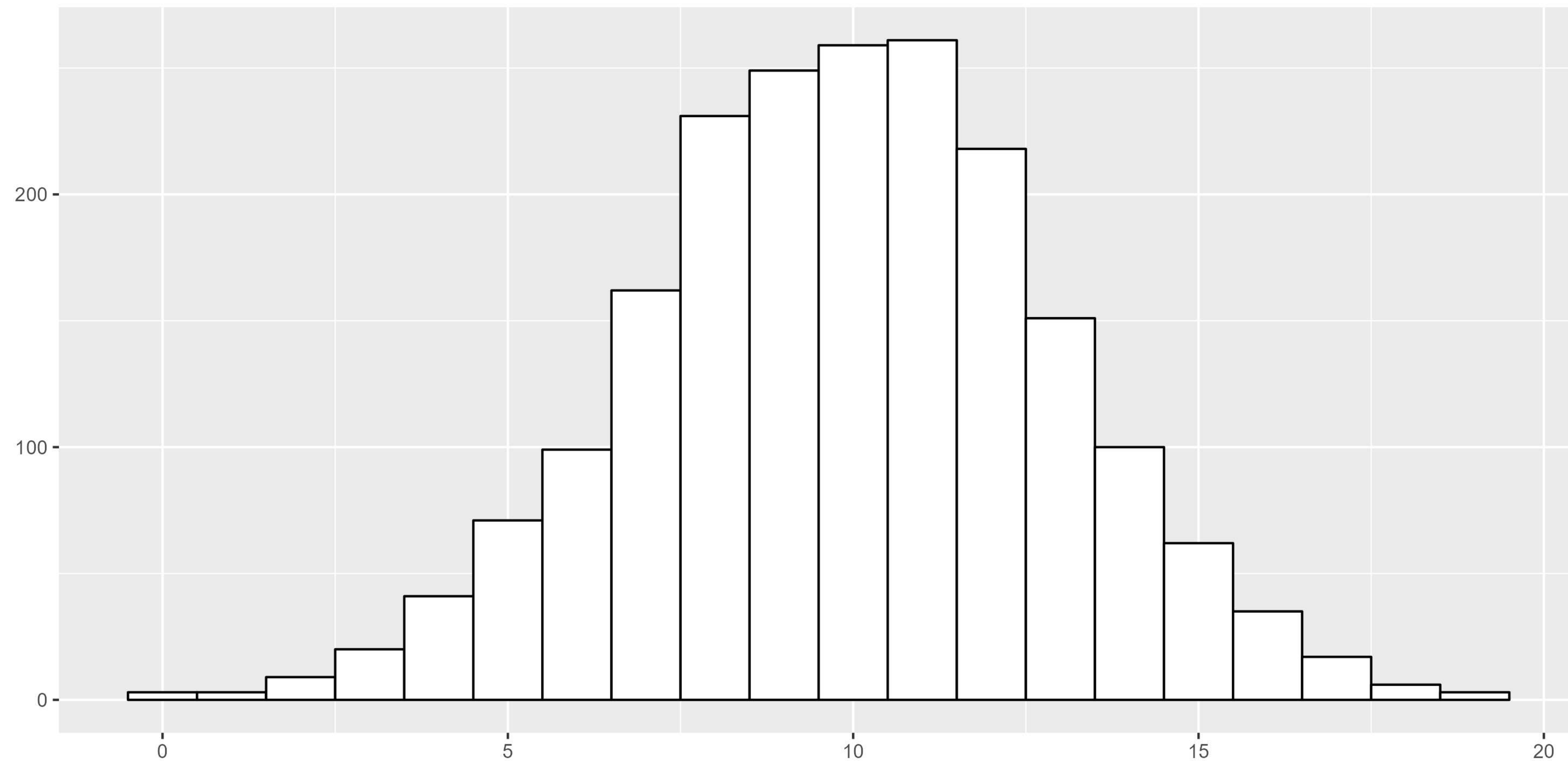


Figure 18: Bell-shaped histogram

Speaker notes

Speaker notes

Here’s a histogram that shows a symmetric distribution, with the most frequent values in the center and frequencies that taper off on either side. This is a bell shaped curve.

Plus or minus one standard deviation

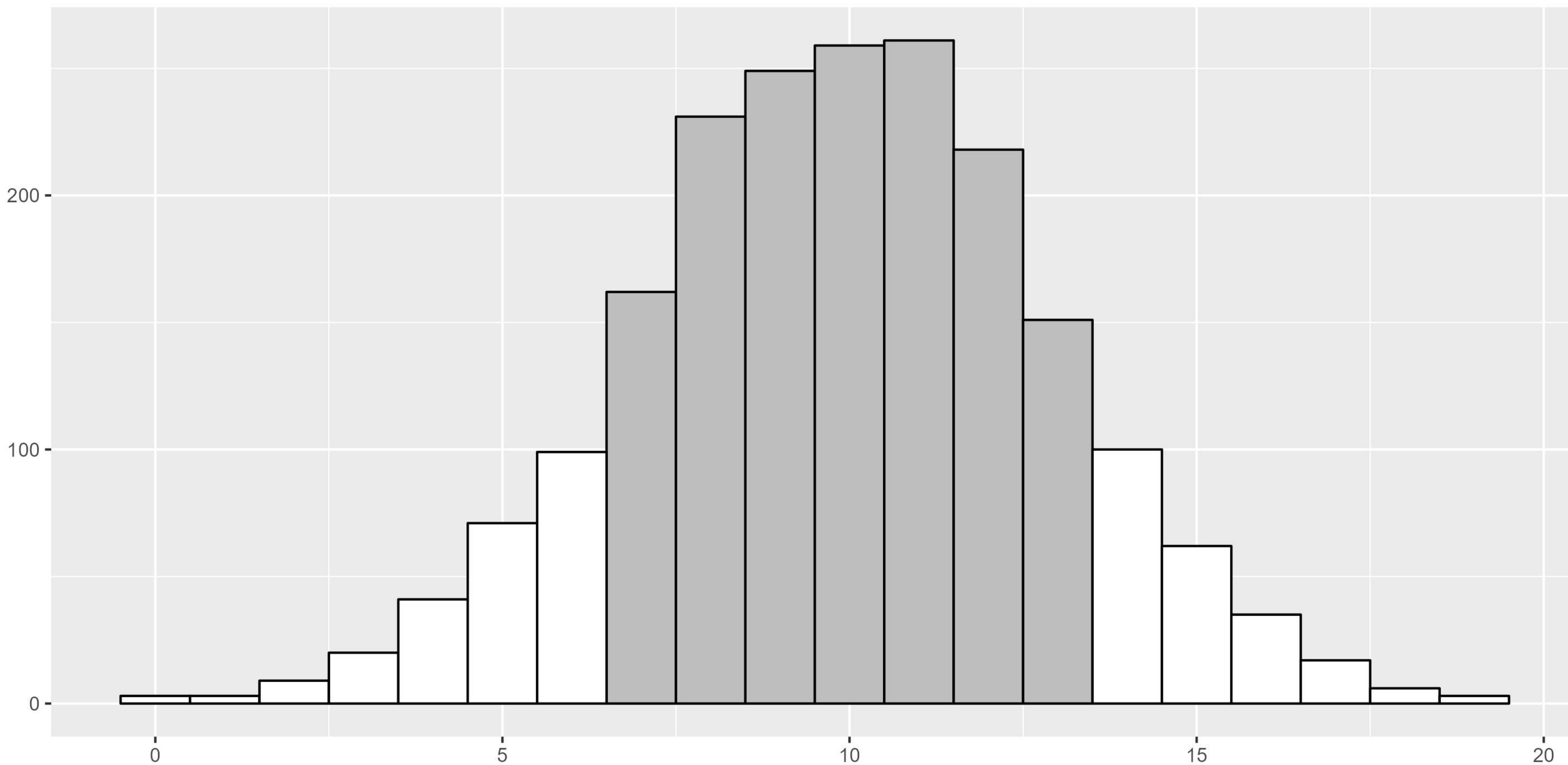


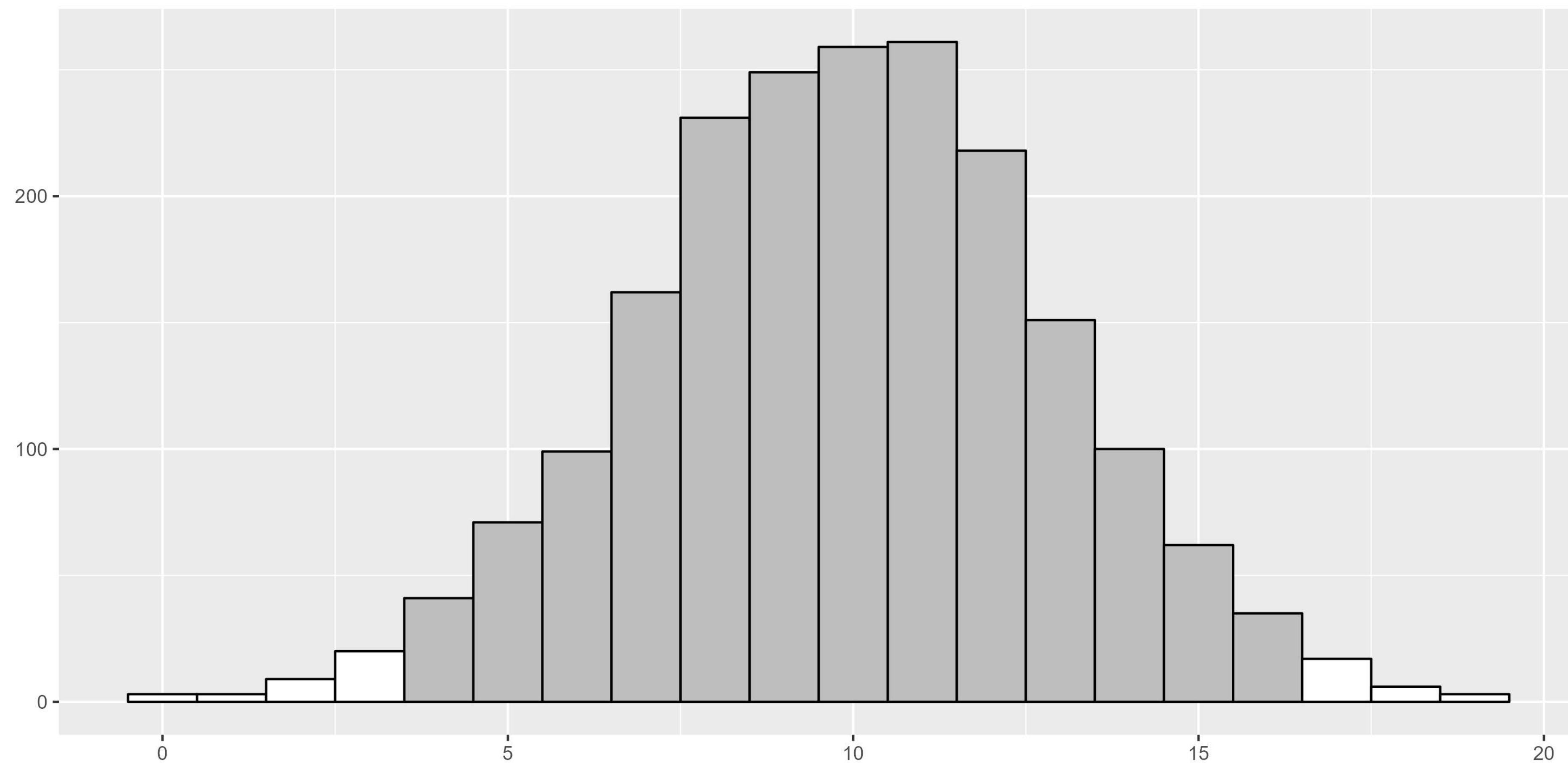
Figure 19: Percentage within one s

Speaker notes

Speaker notes

This shows the bell shaped curve with the data within one standard deviation of the mean highlighted in gray. Roughly 68% of the data lies within one standard deviation of the mean. This is only true if the variation follows a bell shaped curve.

Plus or minus two standard deviations



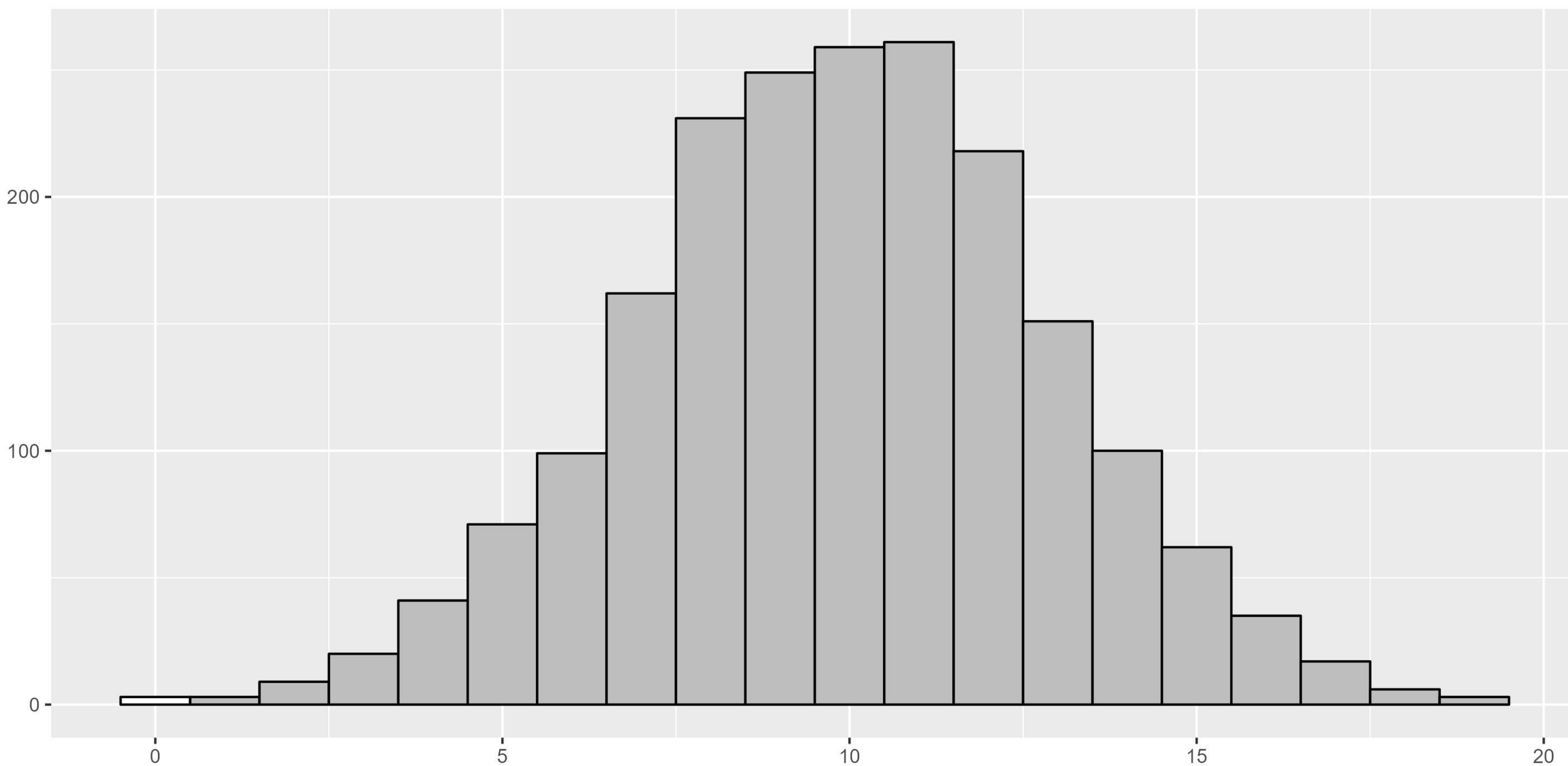
74

Speaker notes

Speaker notes

This shows the bell shaped curve with the data within two standard deviations of the mean highlighted in gray. Roughly 95% of the data lies within one standard deviation of the mean. This is only true if the variation follows a bell shaped curve.

Plus or minus three standard deviations



Speaker notes

Speaker notes

This shows the bell shaped curve with the data within two standard deviations of the mean highlighted in gray. Roughly 95% of the data lies within one standard deviation of the mean. This is only true if the variation follows a bell shaped curve.

Lin 2022, PMID: 36126916

> Prim Care Companion CNS Disord. 2022 Sep 13;24(5):21m03173. doi: 10.4088/PCC.21m03173.

Unemployment, Homelessness, and Other Societal Outcomes Among US Veterans With Schizophrenia Relapse: A Retrospective Cohort Study

Dee Lin [1] [2], Hyunchung Kim [3], Keiko Wada [3], Maya Aboumrad [4], Ethan Powell [4], Gabrielle Zwain [4], Carmela Benson [1], Aimee M Near [3]

Affiliations + expand

PMID: 36126916 DOI: 10.4088/PCC.21m03173

Figure 22: Lin et al 2022

Speaker notes

Speaker notes

Lin et al 2022 patient ages

Variable	Relapse Cohort (n = 16,862)	Nonrelapse Cohort (n = 16,862)
Age at index, y		
Mean (SD)	56.4 (13.3)	56.6 (13.0)
Median (Q1, Q3)	58 (51, 64)	59 (51, 65)
Minimum	20	20
Maximum	99	98

Figure 23: Excerpt from Table 1 of Lin et al 2022: ages

Speaker notes

Speaker notes

Lin et al 2022 Charlson Comorbidity Index

CCI		
Mean (SD)	2.5 (2.7)	2.4 (2.4)
Median (Q1, Q3)	2 (0, 4)	2 (0, 4)
Minimum	0	0
Maximum	20	18

Figure 24: Excerpt from Table 1 of Lin et al 2022: CCI

Speaker notes

Speaker notes

Lin et al 2022 PHQ-2 scores

PHQ-2 score, n (%)[f]	9,472 (56.1)	10,848 (64.4)
Mean (SD)	1.1 (1.7)	1.0 (1.6)
Median (Q1, Q3)	0 (0, 2)	0 (0, 2)
Minimum	0	0
Maximum	6	6

Figure 25: Excerpt from Table 1 of Lin et al 2022: PHQ-2

Speaker notes

Speaker notes

Tosato 2021, PMID: 34352201

J Am Med Dir Assoc. 2021 Sep;22(9):1840-1844. doi: 10.1016/j.jamda.2021.07.003. Epub 2021 Jul 19.

Prevalence and Predictors of Persistence of COVID-19 Symptoms in Older Adults: A Single-Center Study

Matteo Tosato [1], Angelo Carfì [1], Ilaria Martis [1], Cristina Pais [1], Francesca Ciciarello [1], Elisabetta Rota [1], Marcello Tritto [1], Andrea Salerno [1], Maria Beatrice Zazzara [1], Anna Maria Martone [1], Annamaria Paglionico [1], Luca Petricca [1], Vincenzo Brandi [1], Gennaro Capalbo [1], Anna Picca [1], Riccardo Calvani [2], Emanuele Marzetti [3], Francesco Landi [3]; Gemelli Against COVID-19 Post-Acute Care Team

Affiliations + expand

PMID: 34352201 PMCID: PMC8286874 DOI: 10.1016/j.jamda.2021.07.003

Figure 26: Tosato et al 2021

Speaker notes

Speaker notes

Tosato 2021, PMID: 34352201 (continued)

Symptom persistence weeks after laboratory-confirmed severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) clearance is a relatively common long-term complication of Coronavirus disease 2019 (COVID-19). Little is known about this phenomenon in older adults. The present study aimed at determining the prevalence of persistent symptoms among older COVID-19 survivors and identifying symptom patterns.

Speaker notes

Speaker notes

Tosato 2021, PMID: 34352201 (continued)

The mean age was 73.1 ± 6.2 years (median 72, interquartile range 27), and 63 (38.4%) were women. The average time elapsed from hospital discharge was 76.8 ± 20.3 days (range 25-109 days).

Speaker notes

Speaker notes

Ielapi 2021, PMID: 34968328

> Nurs Rep. 2021 Jul 12;11(3):530-535. doi: 10.3390/nursrep11030050.

Insomnia Prevalence among Italian Night-Shift Nurses

Nicola Ielapi [1] [2], Michele Andreucci [3], Umberto Marcello Bracale [4], Davide Costa [2] [5], Egidio Bevacqua [2] [6], Andrea Bitonti [7], Sabrina Mellace [8], Gianluca Buffone [9], Stefano Candido [10], Michele Provenzano [6], Raffaele Serra [2] [6]

Affiliations + expand

PMID: 34968328 PMCID: PMC8608071 DOI: 10.3390/nursrep11030050

Figure 27: Ielapi et al 2021

Ielapi 2021, PMID: 34968328 (continued)

Background. Insomnia is one of the major health problems related with a decrease in quality of life (QOL) and also in poor functioning in night-shift nurses, that also may negatively affect patients’ care. The aim of this study is to evaluate the prevalence of insomnia in night shift nurses.

Ielapi 2021, PMID: 34968328 (continued)

Excerpt from Table 1.

Data reported as mean ± standard deviation or median [Q1-Q3]

```
Overall (n = 2'355)
Age, years  40.4 ± 10.3
Months of work 168 [72-300]
Night shifts per month, number  6.3 ± 1.4
Time to reach workplace, minutes    45 [45-65]
Rest time, minutes  180 [4-240]
Rest in the afternoon, minutes  30 [0-120]
Number of coffees, mean 2.5 ± 1.5
Number of coffees during night shift, mean  1.4 ± 1.1
```

85

Visualization

Proportions

Bar chart

Error bars

Box plot

Histogram

Plot all the data

Rug plot

Which visualization to choose?

Tables versus graphs

How should you display continuous data

How should you display categorical data

How do you illustrate trends and patterns

What are some common mistakes in the choice of colors

http://www.pmean.com/posts/misuse-of-gradient/

http://blog.pmean.com/rainbows/

Primary colors

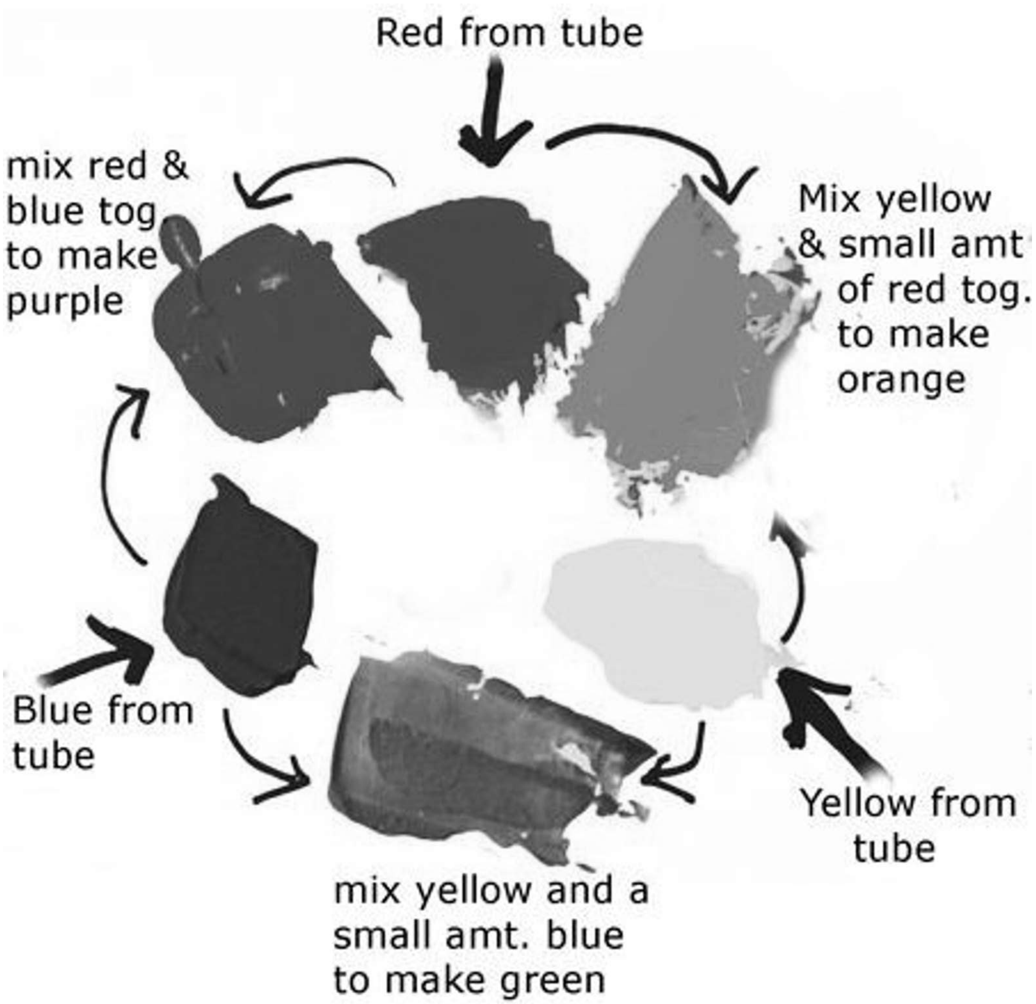


Figure 28: Color combinations

100

The RGB color system

- #rrggbb format
  - #000000 is pure black
  - #FFFFFF is pure white
  - #FF0000 is pure red
  - #00FF00 is pure green
  - #0000FF is pure blue
- You can mix and match to get 16,777,216 colors
  - #800080 is purple, #FF69B4 is pink, #40E0D0 is turquoise

101

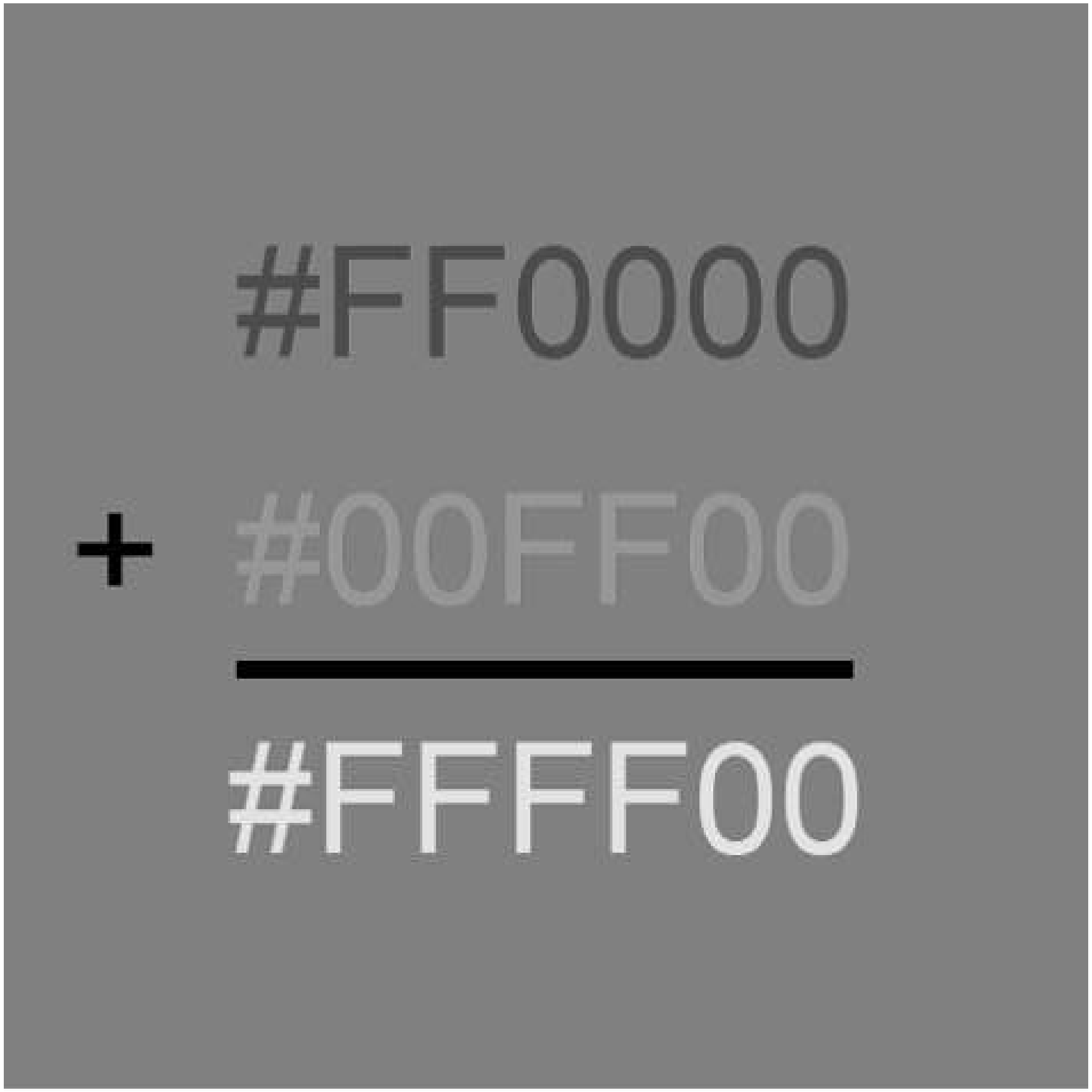


Figure 29: Red plus green equals yellow

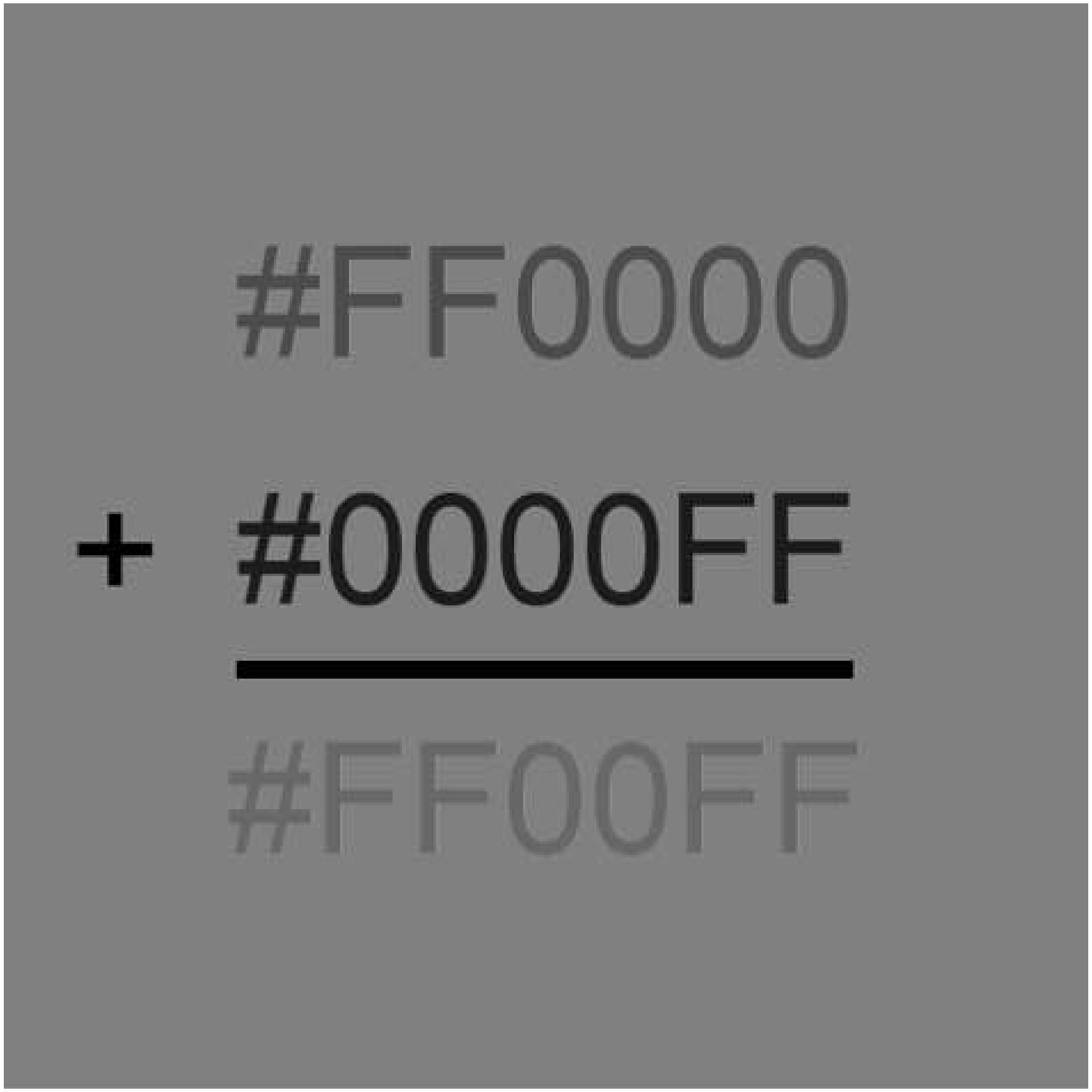


Figure 30: Red plus blue equals magenta

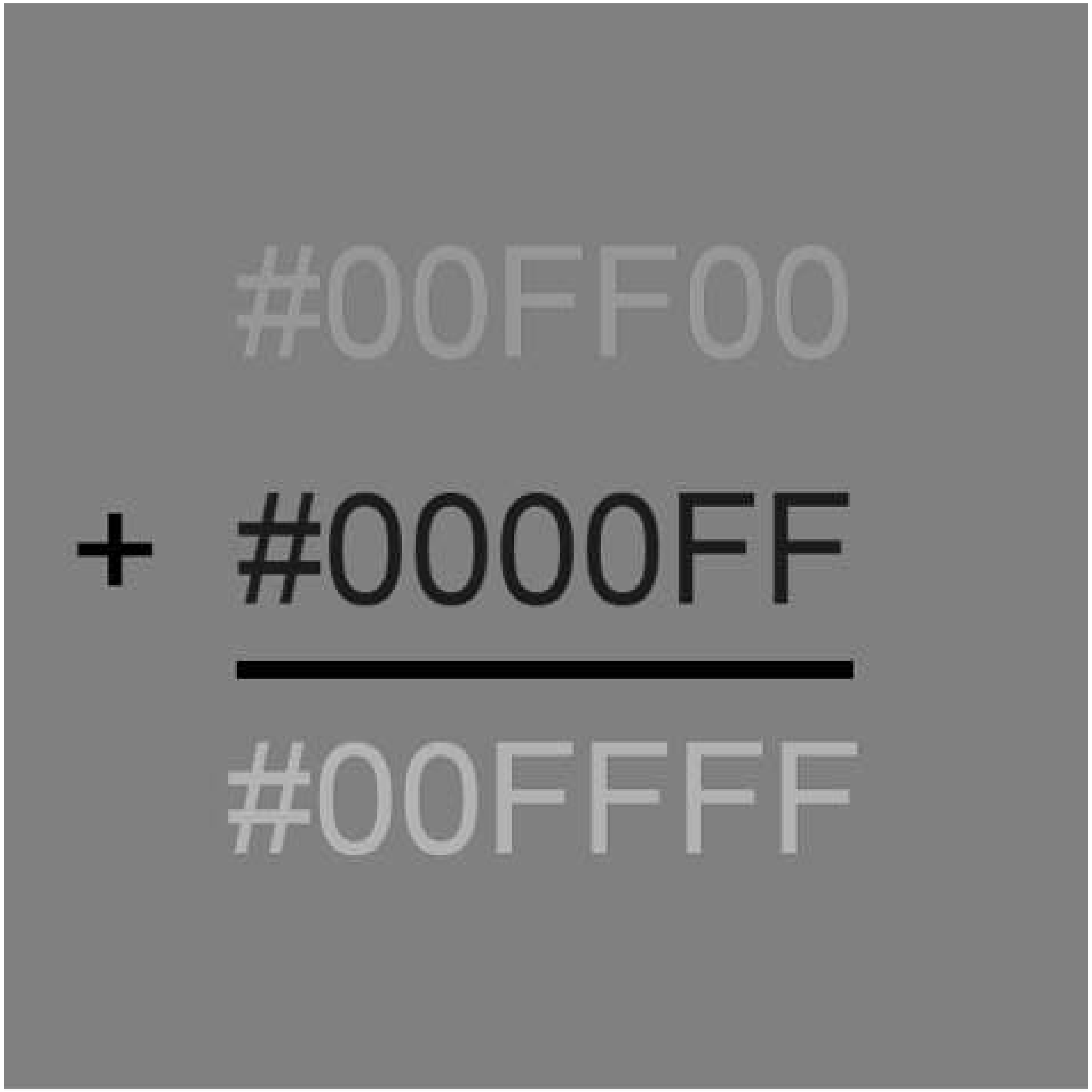


Figure 31: Green plus blue equals cyan

The color cube

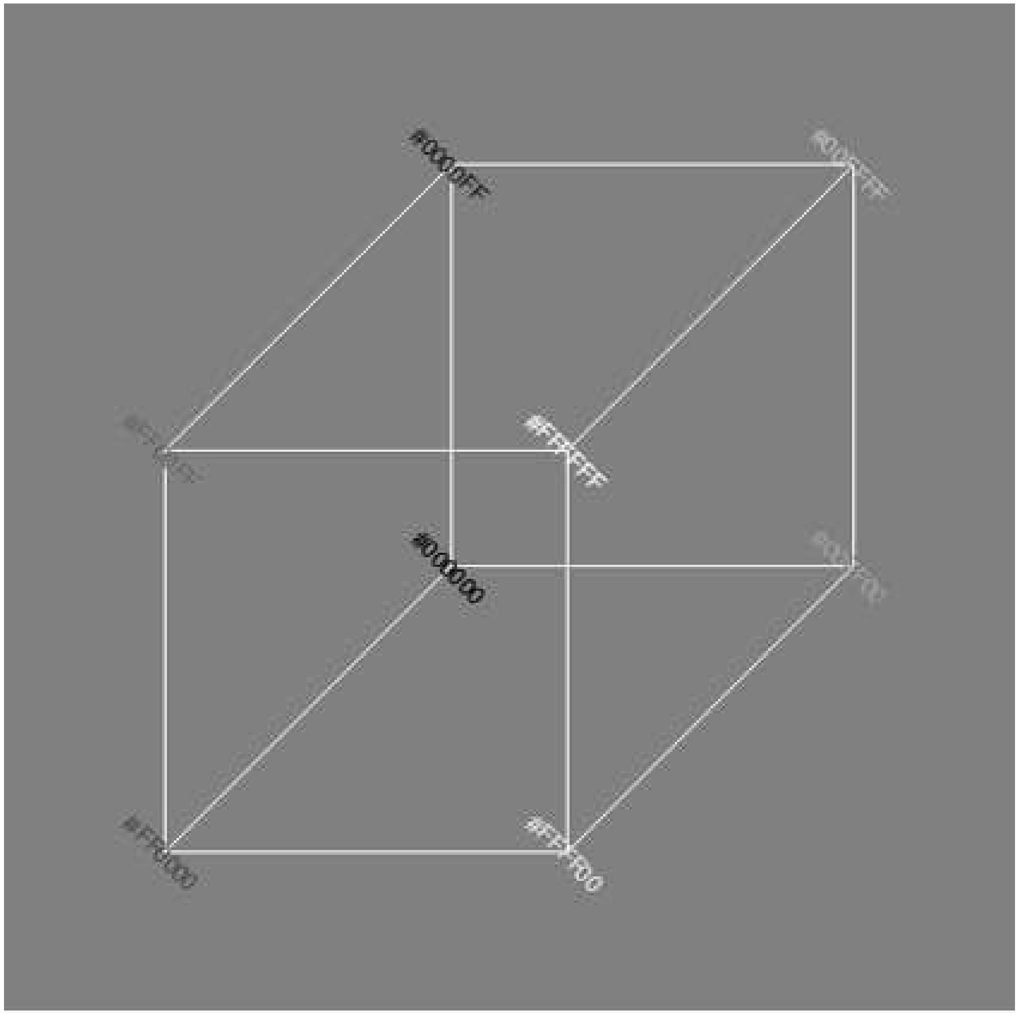


Figure 32: Illustration of the color cube

The color cylinder

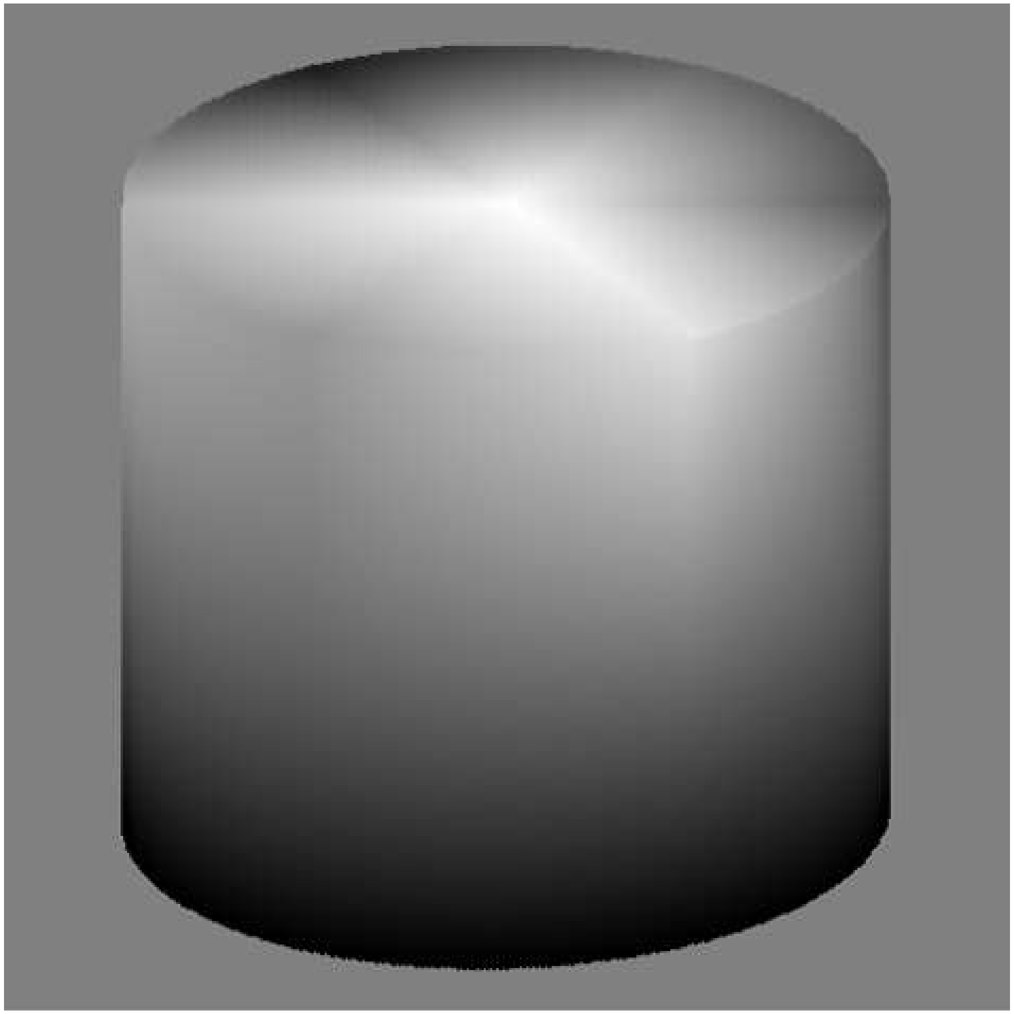
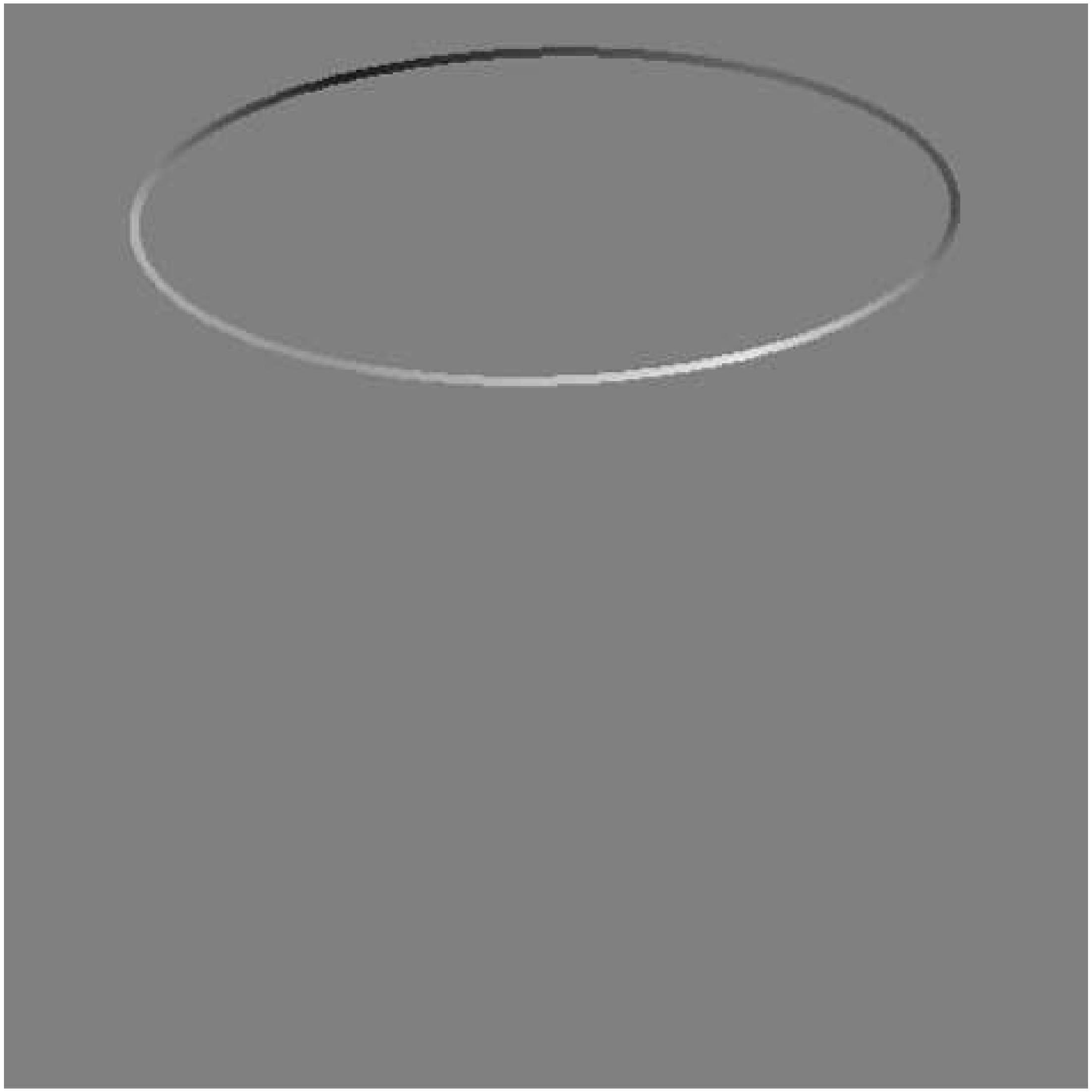


Figure 33: Color cylinder

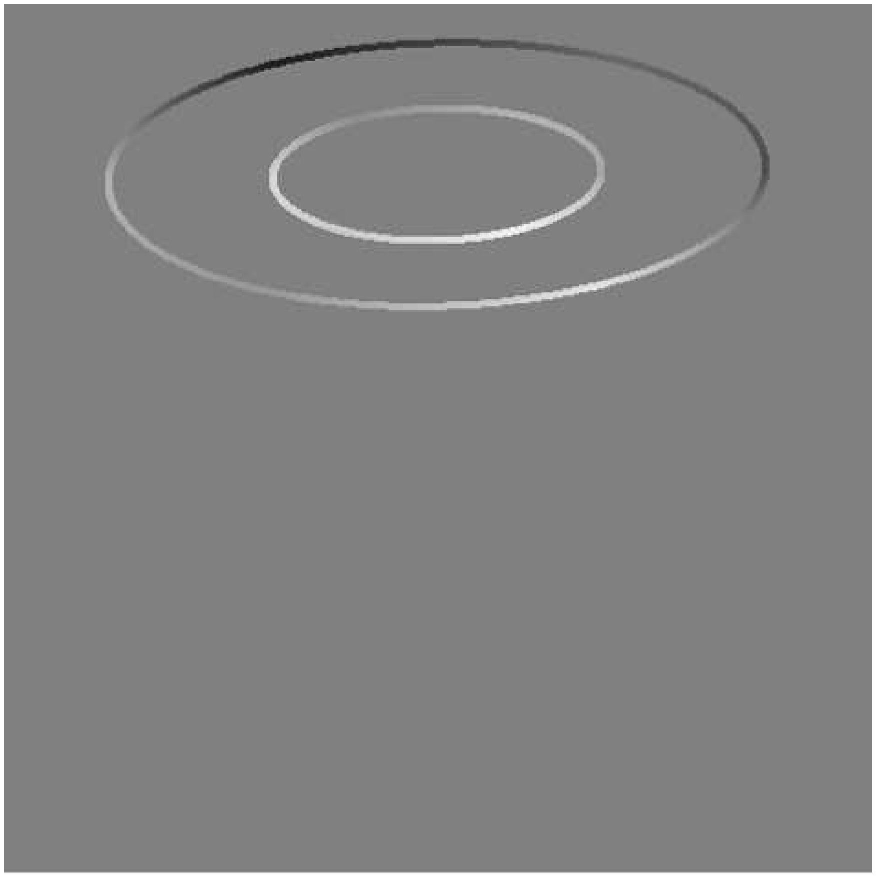
Rainbow



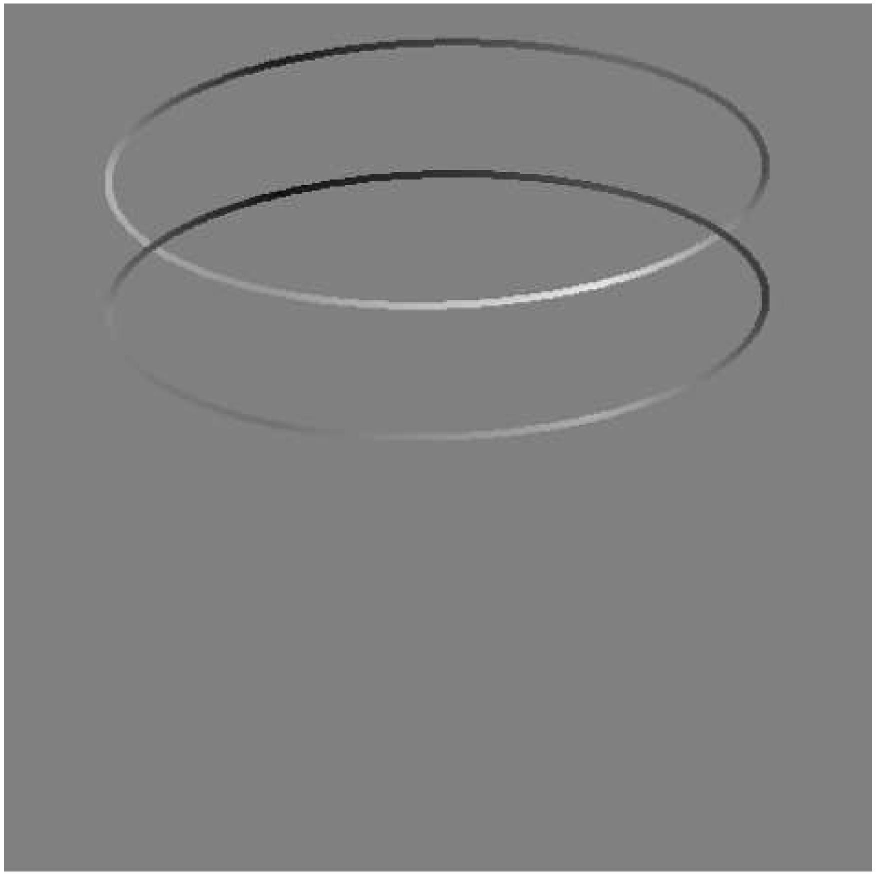
Harsh contrasts

108

Lighter rainbow



Darker rainbow



Gentler contrasts

Equally spaced hues

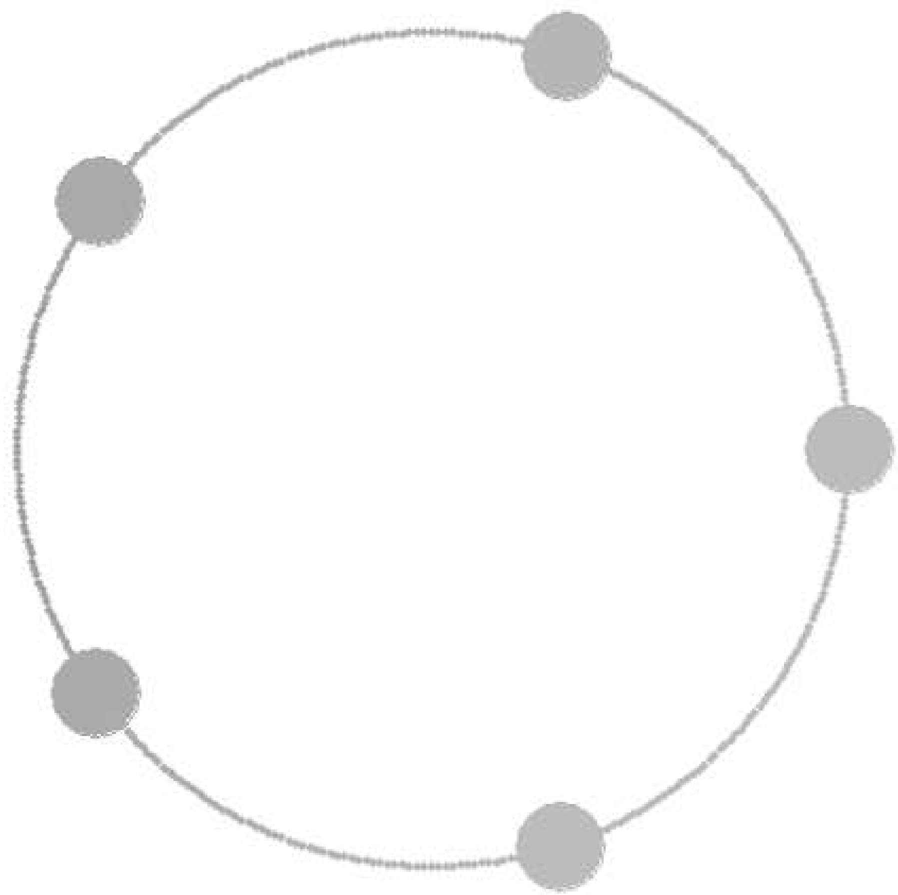


Figure 34: Color choices for nominal data

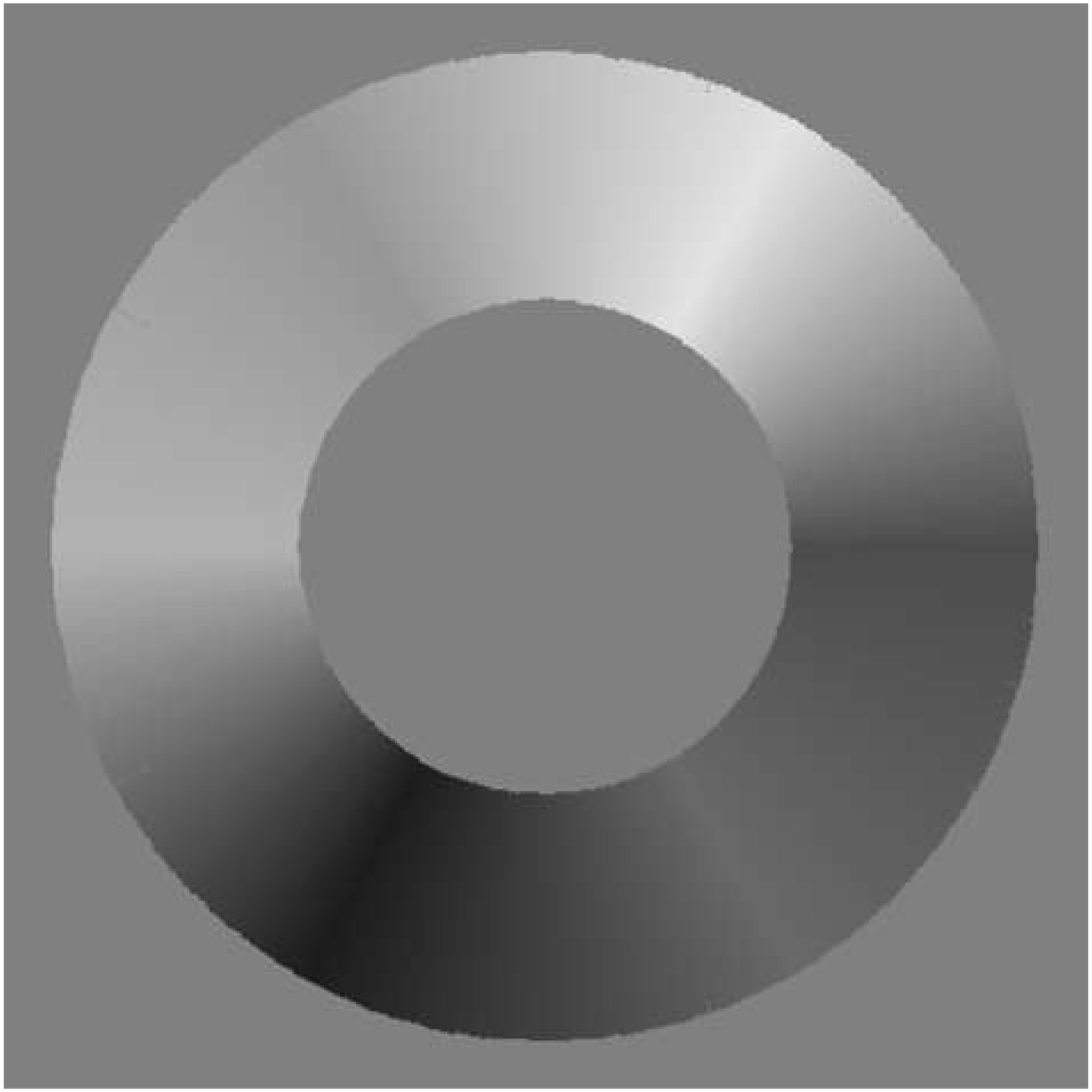


Figure 35: Illustration of the rainbow gradient

#4 TOO MANY COLORS

photo source: pinterest.com

Honestly, we find the best application for this quote in fashion: *"simplicity is the ultimate form of sophistication"*. Keeping it simple might be very difficult for several men, and, looking to the picture above, it really is.

Figure 36: Clothing mistake: using too many colors

ADvertisement with a single red umbrella

Speaker notes

Speaker notes

Graphic designers have known for quite a while that a restrained use of colors can be very effective. Here is an image from a YouTube video clip,

The Travelers - Look under the Umberella commercial (1986). Retrieved 2019-09-07 from https://www.youtube.com/watch?v=3zQX66jd_c0

The single red umbrella in a sea of black umbrellas stands out. Your eye can’t help but follow this umbrella as it travels across the screen from left to right. It’s a very powerful image.

A small dollop of color in your visualizations can be far more effective than using a whole bunch of different colors.

Figure 37: Use of color to highlight a single individual

Speaker notes

Speaker notes

Here is a second example, from the movie, Legally Blonde. In this scene, the main character, Elle Woods, played by Reese Witherspoon, shows her individuality by opening up a bright orange and white Macintosh computer. All the other students are using generic black laptops.

This has practical implications for data visualization.

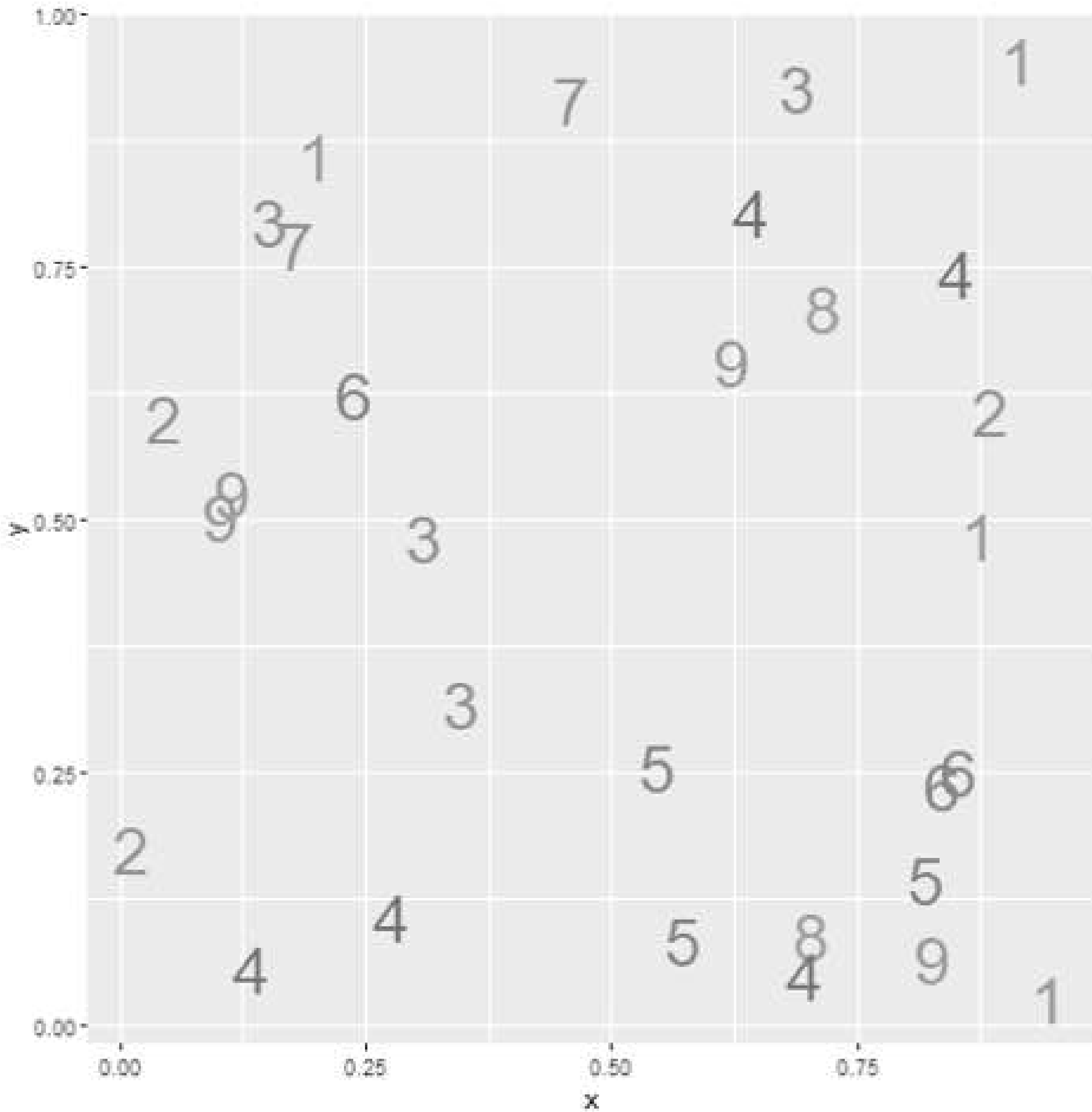


Figure 38: How many “5’s” are in this figure?

Speaker notes

Speaker notes

Here’s a simple exercise, count the number of “5’s” on this graph. Don’t include the “5” that appears in the caption.

When you have an answer, type it in the chat box.

[Pause here]

Now I did try to help by using a different color for each number.

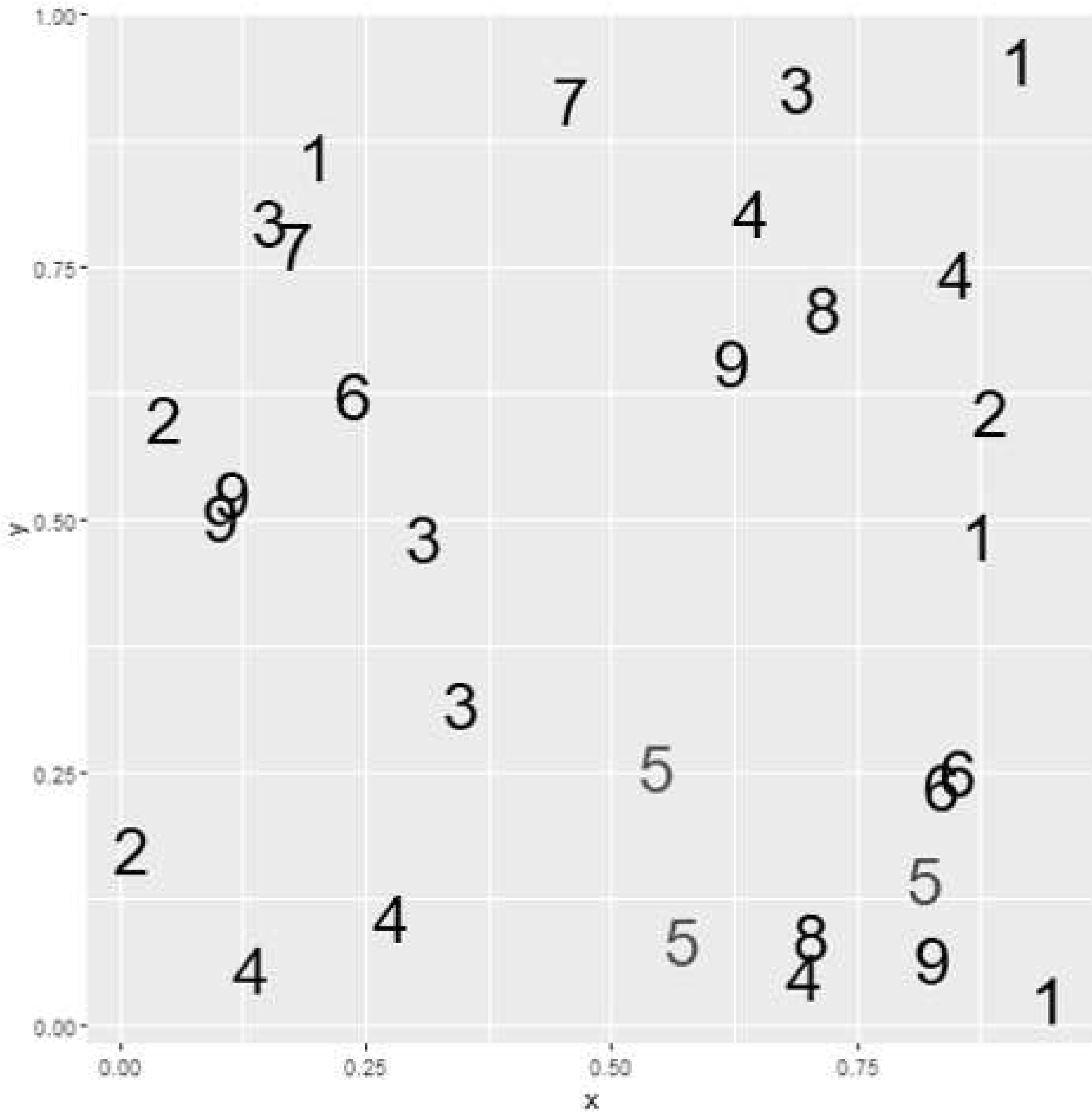


Figure 39: Repeat question. How many “5’s” are in this figure?

117

Speaker notes

Speaker notes

Okay, now repeat this exercise. How many “5’s” do you count? Notice how much faster it is when there is are two colors insead of nine.

Repeat quiz questions