Statistical Faux Pas

Statistical Faux Pas (1)

Facts vs. Hypotheses

"It is a capital mistake to theorize before one has data. Insensibly one begins to twist facts to suit theories, instead of theories to suit facts."

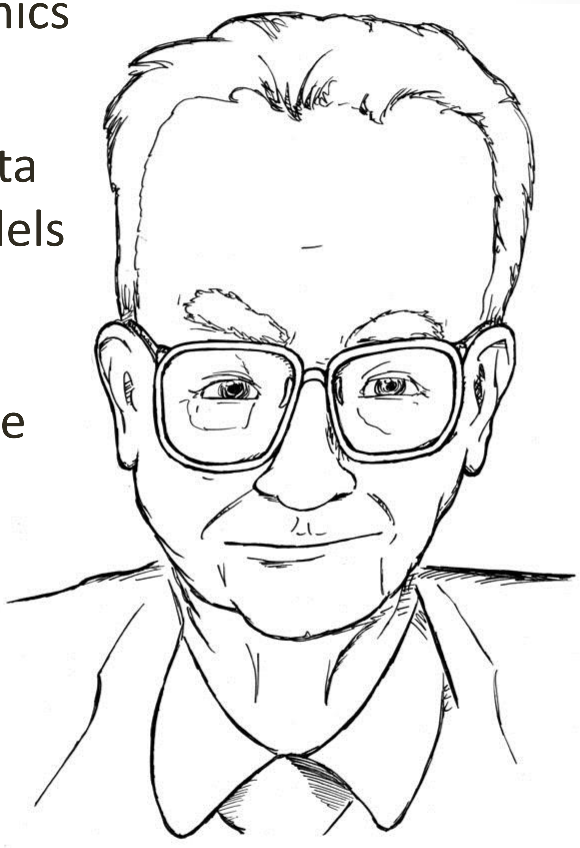
Sir Arthur Conan Doyle as the character of Sherlock Holmes

Statistical Faux Pas (2)

Facts vs. Hypotheses

- "If you torture the data long enough, it will confess," Ronald Coase, Nobel Prize in Economics

- Scientists are not interested in the data for data sake. Scientists want to use data to build models that help us understand nature. Good hypotheses are hard to come by and in my experience, easily one half of all scientists have twisted interpretations of their experimental results to support their pet hypotheses.



Statistical Faux Pas (3)

Facts vs. Hypotheses

In a nursing school we found that if the student's race was "Missing" then the students were more likely to dropout.

At first, we thought that this missing race information indicated that their was an ethnicity that pre-disposed these students to drop out.

But, we could not find any ethnicity that had a significantly higher retention or dropout rate.

In fact, further investigation revealed that the proportion of ethnicities was the same for the overall student population and those students whose race was categorized as "Missing".

Later, we determined that most of the students who filled out the forms themselves did not enter information on their ethnicity. Only those students who were personally assisted by a (diligent) registrar entered a value for race. Further analysis indicated that personal assistance by a registrar, regardless of race, correlated with high retention rate.

Statistical Faux Pas (4)

Facts vs. Hypotheses

- One might conclude: Facts before Hypotheses!

- On the other hand, there are arguments for Hypotheses before Facts:
  - We need hypotheses to guide research. Without a hypothesis we wouldn't know what to data to collect.
  - Hypotheses before Facts prevents cherry picking or shot-gunning of hypotheses until a hypothesis fits.

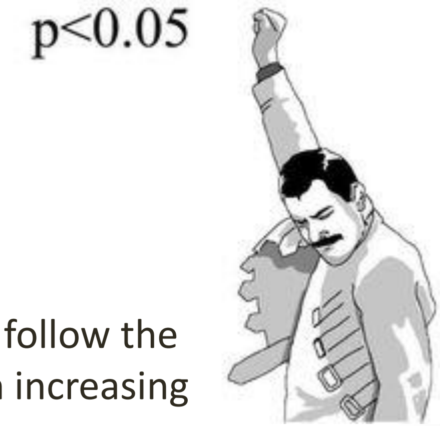
- Could we use Facts before Hypotheses and then use the p-value to determine if a hypothesis is good?

Statistical Faux Pas (5)

Misuse of p-Value

- How is p-value misused?
  - Shot gunning or cherry picking hypotheses
  - Misunderstanding the nature of a p-value.

Statistical Faux Pas (6)



Misuse of p-Value

- How is p-value misused?
  - Do Jelly Beans Cause Acne with $p < 0.05$?
  - http://xkcd.com/882/
  - The null hypothesis states that the observed variations do not follow the hypothesis. If you choose enough hypotheses then there is an increasing chance that you will find a null hypothesis that has a low p-value.
  - Commonly we use a p-value of < 0.05. That means that there is “only” a 5% chance that the null hypothesis is true.
  - If the observed p-value < 0.05, then we might mistakenly assume that there is a 95% chance that the hypothesis accounts for the observations.
  - How many hypotheses (n) should we test if we want a more than even (50%) chance to find 1 or more p-values (p) at less than 5% from random data?
    - $0.5 < 1 - (1-p)^n$; for p = 0.05 we find: n ≥ 14

Statistical Faux Pas (7)

"Data don't make any sense, we will have to resort to statistics."

Misuse of p-Value

How is p-value misused?

Misuse of p-values and experimental design

- After a large, epidemiological study failed to support a hypothesis, the researchers wanted to justify their grant. They looked for any pattern in their data. They transformed their data in as many ways as they could to find a pattern.
- When they found a pattern they retrospectively formulated a hypothesis and then they determined if that hypothesis had a p-value < 0.05, as is common in such studies. A a p-value < 0.05 means that there is only a 5% probability that the null hypothesis accounts for the patterns.
- The researchers announced many (50) hypotheses that were "verified" by this method. Soon colleagues educated them: Constructing a post-facto hypothesis, is similar to re-using training data as testing data.
- Then the researchers randomly partitioned their data into a pattern search dataset and a pattern corroboration dataset. Although, they corroborated 1 of the patterns, this search was still statistically insignificant because we expect that about 5% of the null hypotheses are valid.

Statistical Faux Pas (8)

Misuse of p-Value

- How can we fix the problem?
- Statisticians need to explain the p-value to us.

- The following slides are adapted from: Statisticians Found One Thing They Can Agree On: It's Time To Stop Misusing p-values By Christie Ashwanden http://fivethirtyeight.com/features/statisticians-found-one-thing-they-can-agree-on-its-time-to-stop-misusing-p-values/

Statistical Faux Pas (9)

<u>Misuse of p-Value</u>

- How can we fix the problem?
- Statisticians need to explain the p-value to us.

Definition by Statisticians to layman:
Informally, a p-value is the probability under a specified statistical model that a statistical summary of the data (for example, the sample mean difference between two compared groups) would be equal to or more extreme than its observed value.

Statistical Faux Pas (10)

Misuse of p-Value

- How can we fix the problem?
- Statisticians need to explain the p-value to us.

"That definition is about as clear as mud"

Definition by Statisticians to layman:
Informally, a p-value is the probability under a specified statistical model that a statistical summary of the data (for example, the sample mean difference between two compared groups) would be equal to or more extreme than its observed value.

"... even scientists can't easily explain p-values"

Statistical Faux Pas (11)

Misuse of p-Value

More links:

- http://www.nature.com/news/statisticians-issue-warning-over-misuse-of-p-values-1.19503
- http://www.nature.com/news/how-scientists-fool-themselves-and-how-they-can-stop-1.18517
- http://www.nature.com/news/scientific-method-statistical-errors-1.14700
- http://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.0020124
- http://www.nature.com/news/statisticians-issue-warning-over-misuse-of-p-values-1.19503

Statistical Faux Pas (12)

Observational Studies

- If we have a problem with collecting facts before developing a hypothesis and we have a problem with developing a hypothesis before collecting facts, then we may want to observe and draw conclusions from observational studies.
- In Search of Excellence (1982) one of the most popular business books of all time. Studied 44 successful "excellent" companies.
- 5 years later 65% of the companies did worse than the S&P 500 Index
- The case study method makes for good drama / story-telling, but it's not good science. Small sample sizes, subject to numerous biases (survivorship, extreme cases,…)
- Other examples: One-off medical studies,

Statistical Faux Pas (13)

Observational Studies

- "Any claim coming from an observational study is most likely to be wrong" (Stanley Young)
- Young and Karr looked at 52 similar published epidemiological findings that were followed by a clinical trial testing the result.
- NONE of the 52 claims replicated in the clinical trials! (5 were significant in the opposite direction.)

Statistical Faux Pas (14)

Observational Studies

- Wrong results from an observational study could be
  - Innocent
  - Not so innocent – sort through the data to find evidence to prove your case and ignore all the other signals
- How to determine if it's a real insight?
  - Test it – conduct a valid experiment to see if the presumed cause and effect relationship holds (e.g. clinical trial, design of experiments (DOE))
  - Get additional, independent data sets and see if the relationship is still present
  - Caution: Most analyses only validate the presence of a relationship. Most analyses do not even show that a cause and effect relationship exist. And, even if a cause and effect relationship exists, we might know which is cause and which is effect.
  - Never allow the same dataset to suggest a relationship AND validate it.

Statistical Faux Pas (15)

Reliance on Descriptive Measures

- We need to understand our data better before we make conclusions.
- We can use descriptive measures to help us understand our data

Statistical Faux Pas (16)

Reliance on Descriptive Measures

Francis determined 6 measures of a dataset.

Property	Value
mean(y)	9
var(x)	11
mean(y)	7.50
var(y)	4.125
cor(x,y)	0.816
lm(y~x)	y = 3 + 0.5x

Statistical Faux Pas (17)

Reliance on Descriptive Measures

What does the dataset look like?
Do the data have outliers?
Do the data form a linear relationship?
Can he extrapolate from this relationship?

Francis determined 6 measures of a dataset.

Property	Value
mean(y)	9
var(x)	11
mean(y)	7.50
var(y)	4.125
cor(x,y)	0.816
lm(y~x)	y = 3 + 0.5x

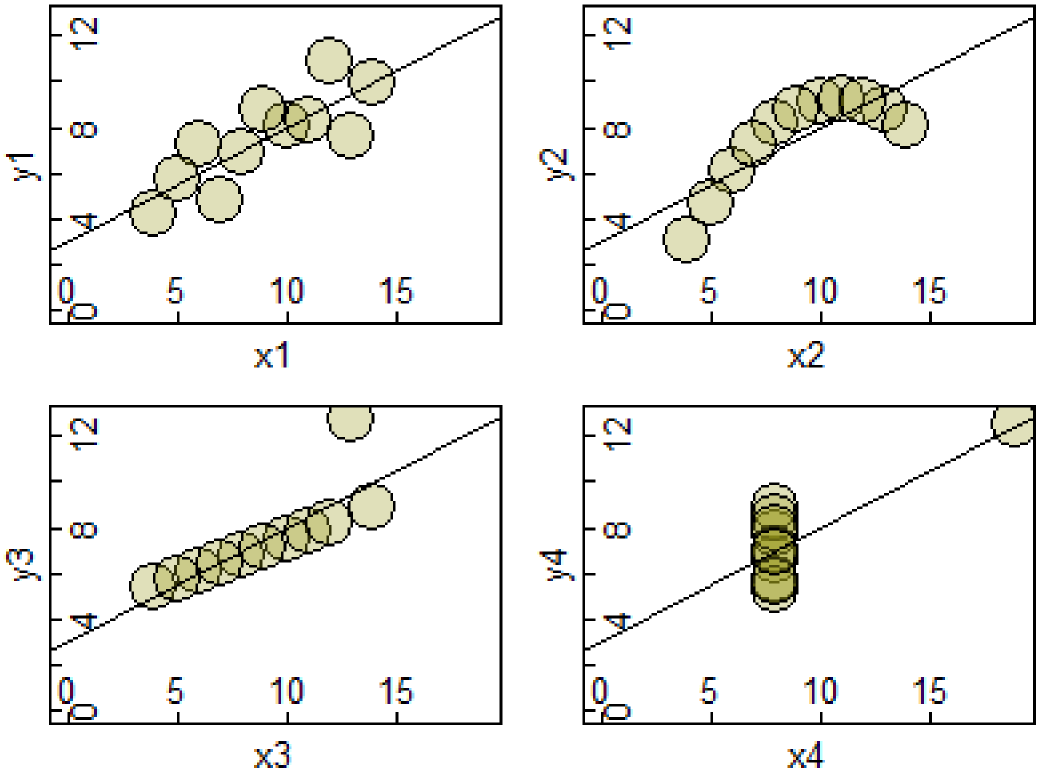
Statistical Faux Pas (18)

Reliance on Descriptive Measures

What does the dataset look like?
Do the data have outliers?
Do the data form a linear relationship?
Can he extrapolate from this relationship?

Francis determined 6 measures of a dataset.

Property	Value
mean(y)	9
var(x)	11
mean(y)	7.50
var(y)	4.125
cor(x,y)	0.816
lm(y~x)	y = 3 + 0.5x



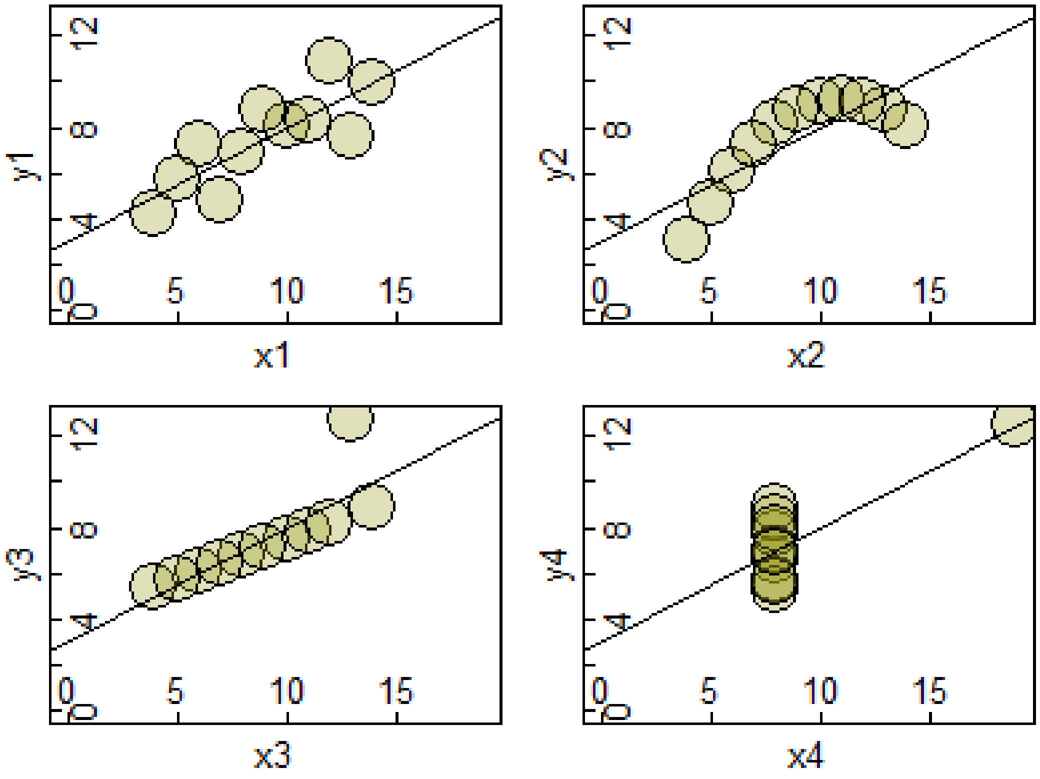
Statistical Faux Pas (19)

Reliance on Descriptive Measures

What does the dataset look like?
Do the data have outliers?
Do the data form a linear relationship?
Can he extrapolate from this relationship?

Francis determined 6 measures of a dataset.

Property	Value
mean(y)	9
var(x)	11
mean(y)	7.50
var(y)	4.125
cor(x,y)	0.816
lm(y~x)	y = 3 + 0.5x



Which data set belongs to these measurements?

Statistical Faux Pas (20)

Reliance on Descriptive Measures

Francis determined 6 measures of a dataset.

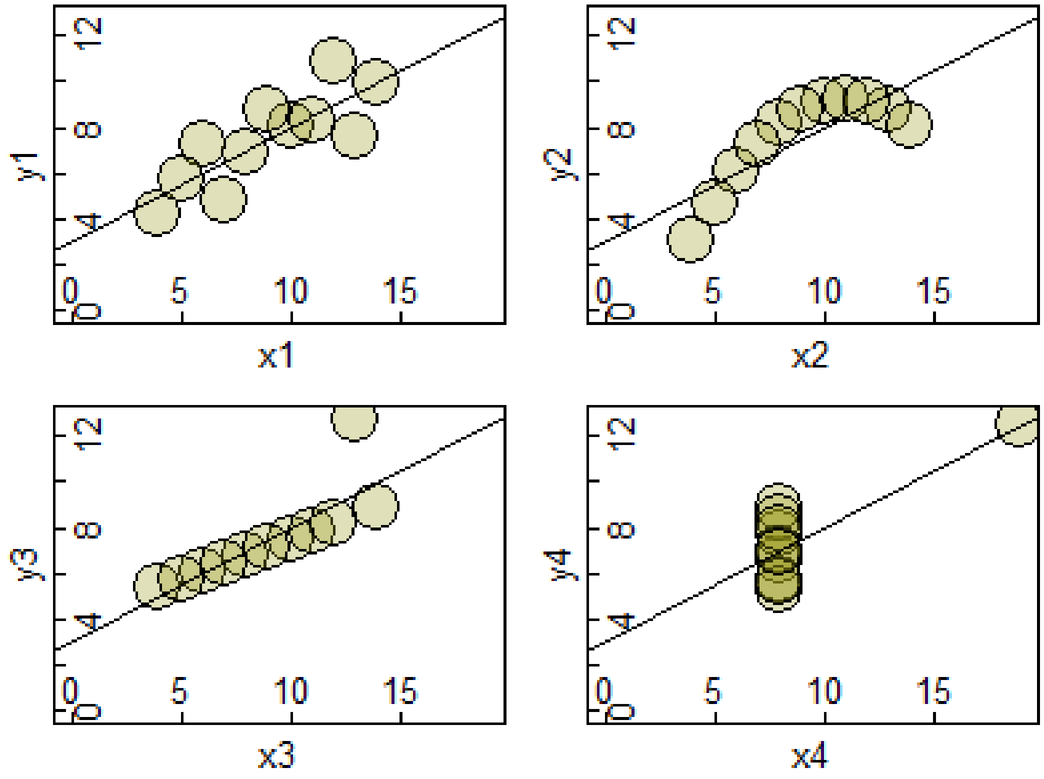
Property	Value
mean(y)	9
var(x)	11
mean(y)	7.50
var(y)	4.125
cor(x,y)	0.816
lm(y~x)	y = 3 + 0.5x

All these data sets have these measurements!

Anscombe's Quartet

See: AnscombeQuartet.R

https://en.wikipedia.org/wiki/Anscombe%27s_quartet



Statistical Faux Pas (21)

- Other problems may arise due to expectations of future performance.

Statistical Faux Pas (22)

Expectations on Performance

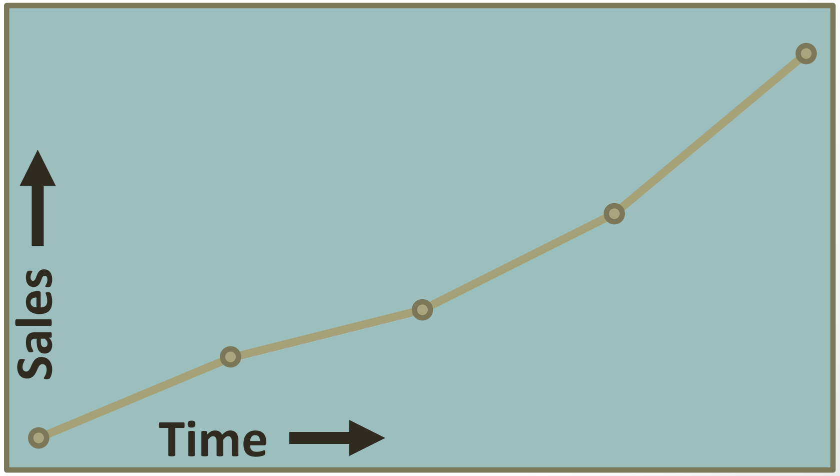
- Management complains: “Our top 1000 customers in 2014 bought 20% less in 2015”
- Management assumes that these customers were disappointed. But, a reduction is expected. The phenomenon is known as “regression to the mean”
- If a measurement of a variable is observed to be extreme, and there is no trend, it will tend to be closer to the average on the next measurement
- Examples:
  - Performance Reviews
  - Sales by Account Managers
  - Sports Illustrated Jinx

- (“Regression to the mean” is the origin of the word regression as in linear regression.)

Statistical Faux Pas (23)

<u>Interpreting Recent Trends</u>

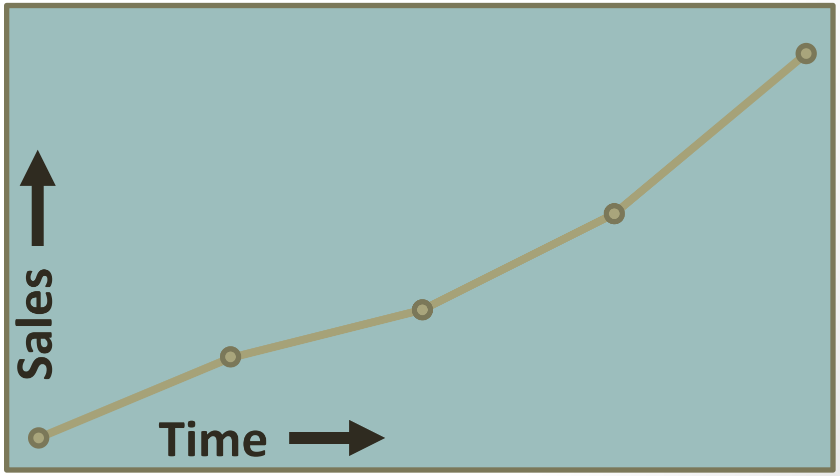
- Sales have increased for the past four months in a row. Are we on a meaningful trend? What are the chances?



Statistical Faux Pas (24)

Interpreting Recent Trends

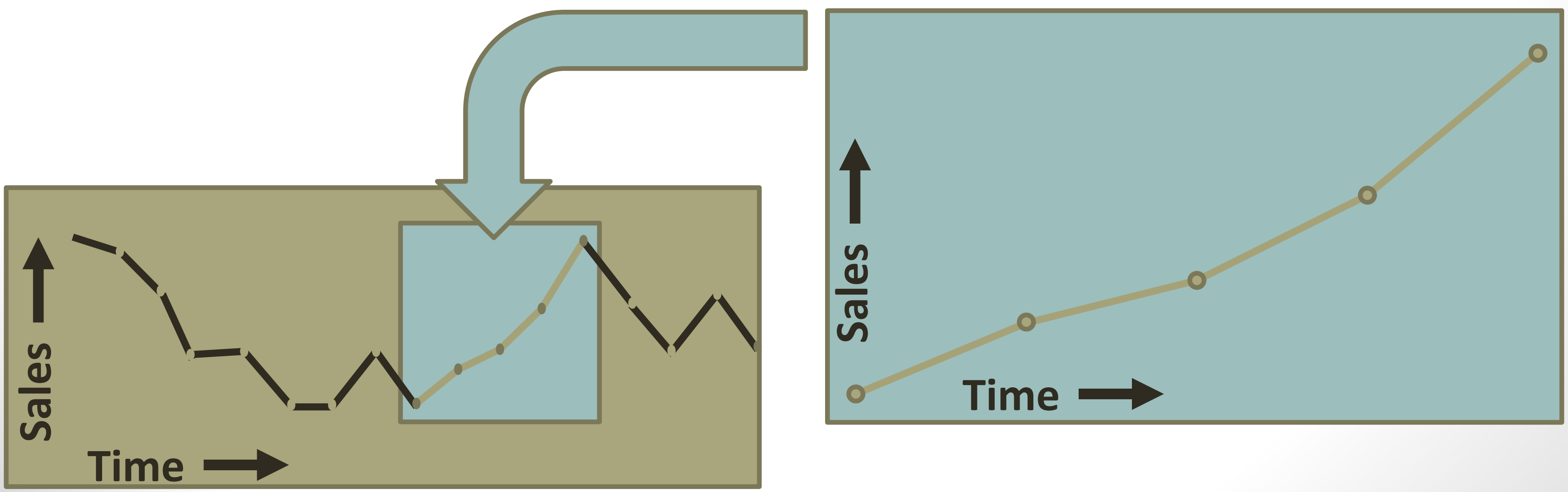
- Sales have increased for the past four months in a row. Are we on a meaningful trend? What are the chances?
- 5 monthly measurements where each successive measurement increased:
- ½ * ½ * ½ * ½ = 6%



Statistical Faux Pas (25)

Interpreting Recent Trends

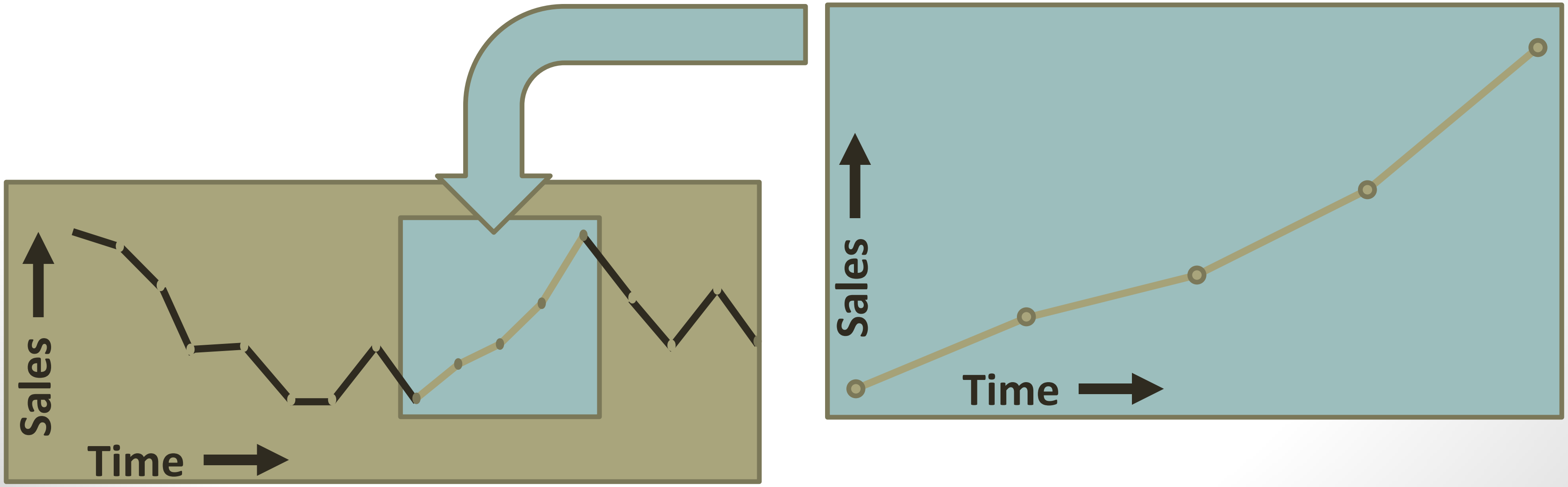
- Sales have increased for the past four months in a row. Are we on a meaningful trend? What are the chances in a year?



Statistical Faux Pas (26)

Interpreting Recent Trends

- Sales have increased or decreased for the past four months in a row. Are we on a meaningful trend? What are the chances in a year?
  - 9 sequences of 4 changes: 0-4, 1-5, …, 8-12
  - 4 sequential increases
  - $1 - (1 - 2^{-4})^9 = 0.44$
- Assume a monthly random measurement. In a year there is a 44% chance of 4 sequential increases.



Statistical Faux Pas (27)

Correlation *vs.* Causation

- A common human trait is to observe two things occurring together and assume one is causing the other
- Examples:
  - Leading Economic Indicators
  - Bad Breath and Heart Disease

- An observed (statistically significant) relationship may be due to
  - Happenstance (i.e. chance or co-incidence)
  - Statistical significance helps, but among 100 relationships with p=0.05, odds are that about 5 will be by chance.
  - Common hidden factor
  - True cause-effect relationship but which direction?

Statistical Faux Pas (28)

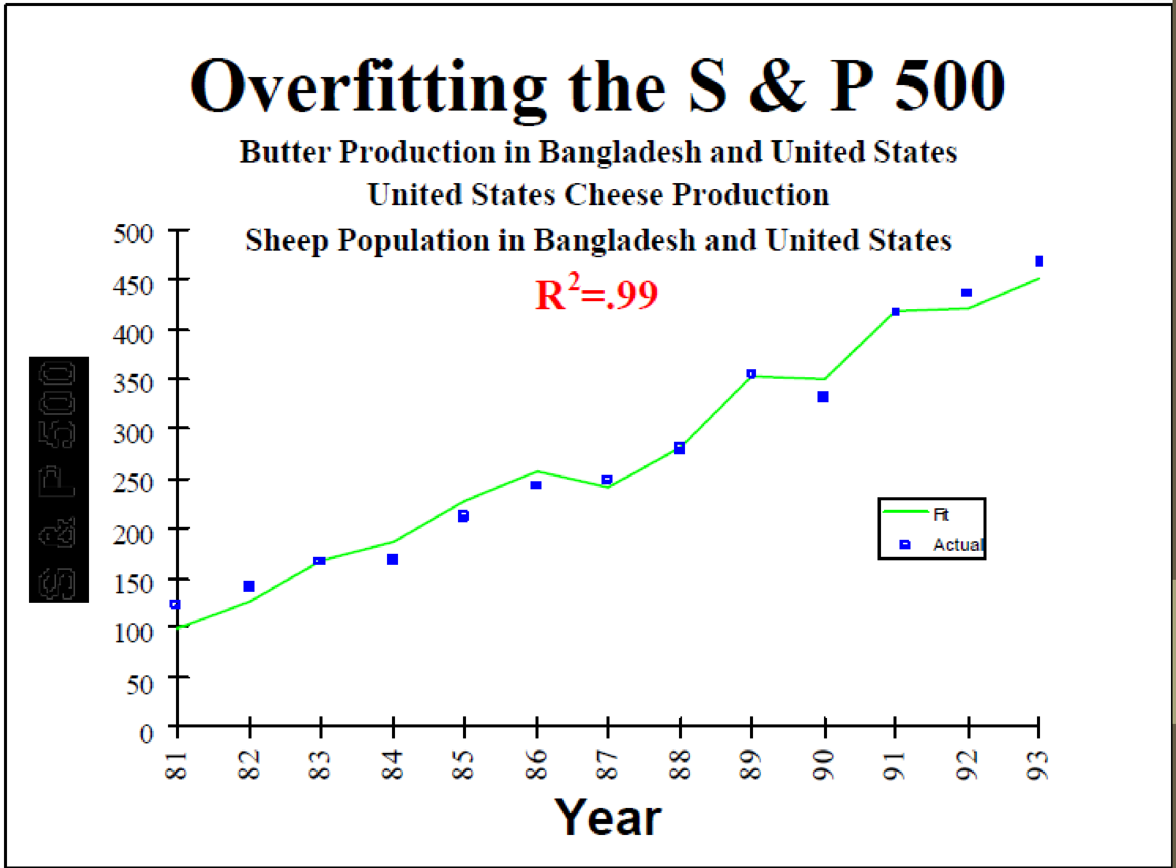
Spurious Relationship

https://en.wikipedia.org/wiki/Spurious_relationship

Statistical Faux Pas (29)

<u>Spurious Relationship</u>

Exact prediction of S&P 500 returns by Ivan O. Kitov, Oleg I. Kitov (See: SSRN-id1045281.pdf)



Statistical Faux Pas (30)

Spurious Relationship

- Redskins Rule (http://en.wikipedia.org/wiki/Redskins_Rule)
    - http://abbottanalytics.blogspot.com/2012/11/why-predictive-modelers-should-be.html

"Our algorithms have linked funny cat videos, UFO reports and searches for tofu pizza. We're now on alert about a suspicious group of cat aliens who infiltrated our pizza industry."

Statistical Faux Pas (31)

Hidden Proxies

- We were using predictive analytics to look for causes of dropouts in a nursing school.
- At one point we looked for professors who were associated with high dropouts or high retention.
- We found one professor whose students had a 100% retention rate. We thought that this result was significant.
- It turned out that this professor had the final class in this two-year program. In other words, drop-outs occurred prior to this professor's class. In fact her class was a pro-seminar and all the students for this class had essentially already graduated.

Statistical Faux Pas (32)

Hidden Proxies

- Proxies and Audience Gullibility:
- Scam artists use proxy attributes in their “predictions”
- A true story from about 20 years ago:
  - A fortune teller went on a radio talk show on KGO in the Bay Area.
  - He demonstrated how he could mimic psychic abilities by getting people to divulge information without their knowledge.
  - After the show, this confessed scam artist was flooded with requests for psychic readings.
  - The audience preferred to believe in his psychic powers and not his confessions.

Statistical Faux Pas (33)

Selective Presentation of Outcomes

- In the 1950's, a convict in Italy, wrote to 80 stockbrokers from prison. He claimed to have insider information from a fellow convict who had been an executive at a local company.
- To 40 stockbrokers he wrote that the stock price would rise in the next two days. To the other 40 stockbrokers he wrote that the stock price would fall.
- After two days he followed up letters to the 40 stockbrokers who received the correct prediction. To half of those he wrote that the stock price would rise and to the other half he wrote that the stock price would fall.
- The prisoner repeated this pattern three more times and then requested a fee from the stockbrokers for additional predictions.

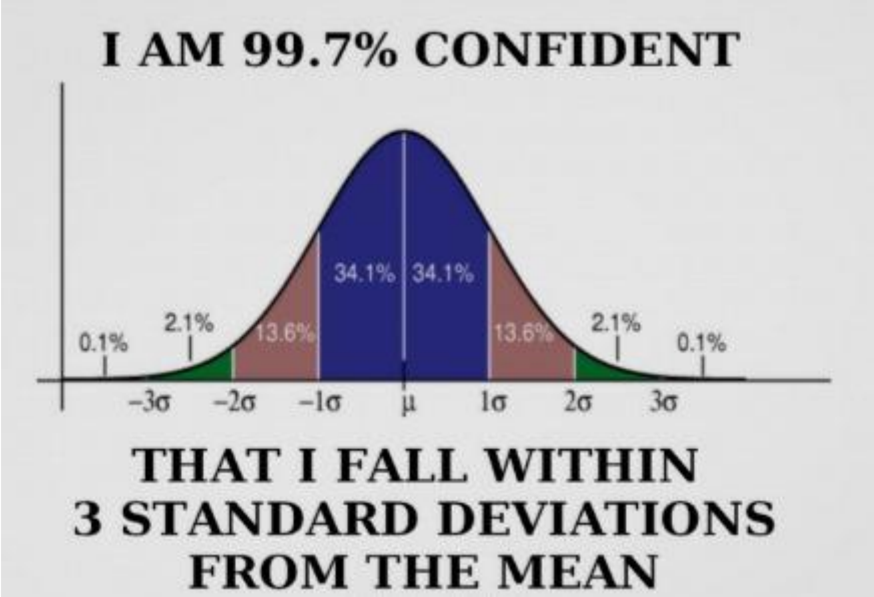
Statistical Faux Pas (34)

Superstition

- HBR Superstitious Learning
- "Superstitious learning takes place when the connection between the cause of an action and the outcomes experienced aren't clear, or are misattributed."
- Some Causes:
  - Expecting high/low performance to remain at that level
  - Interpreting trends that could be due to randomness
  - One-off occurrences
  - Causation inferred from correlation

Statistical Faux Pas (35)

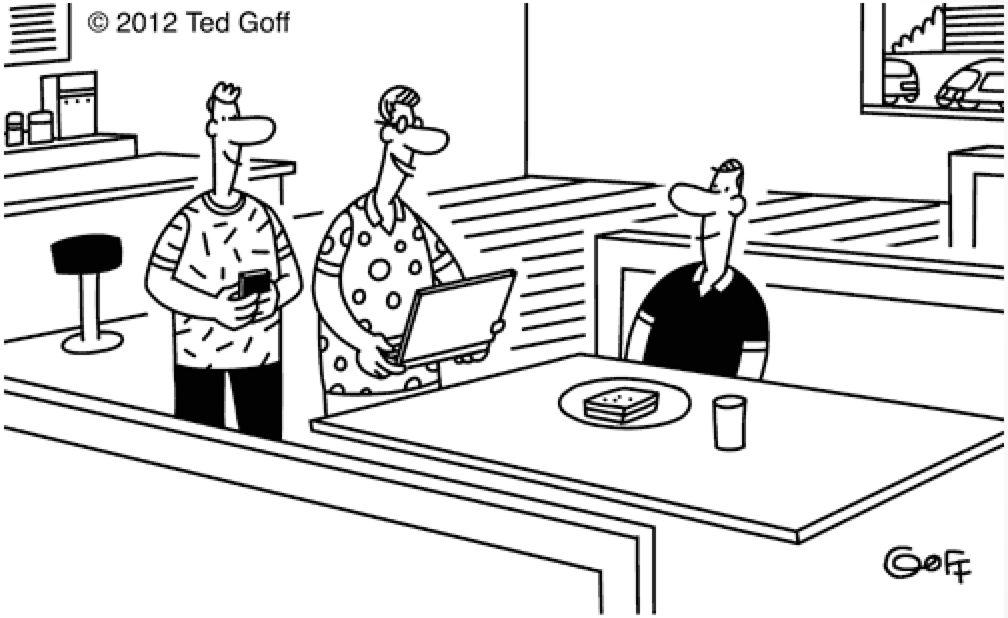
Tautology



Statistical Faux Pas (36)

- Some more links
  - http://skepdic.com/perfectprediction.html
  - http://www.investorhome.com/scam.htm
  - http://www.forbes.com/sites/davidleinweber/2012/07/24/stupid-data-miner-tricks-quants-fooling-themselves-the-economic-indicator-in-your-pants/
  - Leo Breiman, Statistical Modeling: The Two Cultures, Statistical Science, 2001, Vol. 16, No. 3, 199–231 http://projecteuclid.org/download/pdf_1/euclid.ss/1009213726

Statistical Faux Pas (37)

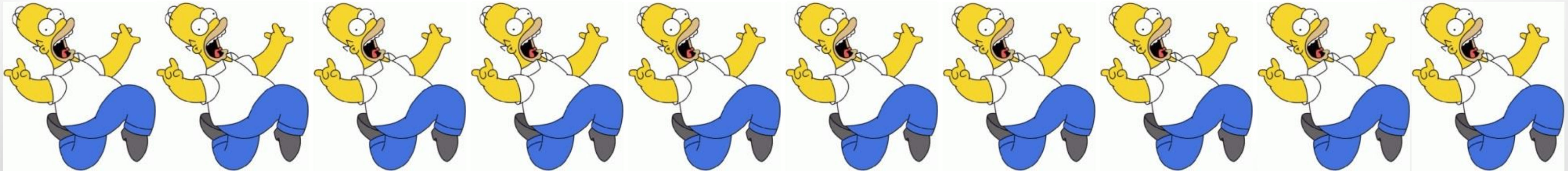


"Twitter and Facebook can't predict the election, but they did predict what you're going to have for lunch: a tuna salad sandwich. You're having the wrong sandwich."

Statistical Faux Pas (38)

Simpson's Paradox

- A trend appears in different groups of data but disappears or reverses when these groups are combined.
- Simpson, E.H. (1951) The interpretation of interaction in contigency tables. Journal of the the Royal Statistical Society, Ser B, 13, 238-241
- Also called the Yule–Simpson effect
- https://en.wikipedia.org/wiki/Simpson's_paradox
- https://www.scientificamerican.com/article/mathematical-games-1976-03/
- http://www.mortalityresearch.com/images/uploads/entry_image/Simpsons_paradox_in_MLB.pdf
- https://www.jstor.org/stable/2984065?seq=1#page_scan_tab_contents
- https://www.jstor.org/stable/2284382?seq=1#page_scan_tab_contents
- See: SimpsonsParadox.R



Statistical Faux Pas (39)

Simpson's Paradox

Player	1995 Hits/ At Bat	1995 BA
Derek Jeter	12/48	0.25
David Justice	104/411	0.253

Higher Batting Average (BA)

Statistical Faux Pas (40)

Simpson's Paradox

Player	1995 Hits/ At Bat	1995 BA
Derek Jeter	12/48	0.25
David Justice	104/411	0.253

Higher Batting Average (BA)

Player	Comb Hits/ At Bat	Comb BA
Derek Jeter	12/48	0.25
David Justice	104/411	0.253

Higher Batting Average (BA)

Statistical Faux Pas (41)

Simpson's Paradox

Player	1995 Hits/ At Bat	1995 BA	1996 Hits/ At Bat	1996 BA
Derek Jeter	12/48	0.25	183/582	0.314
David Justice	104/411	0.253	45/140	0.321

Higher Batting Average (BA)

Statistical Faux Pas (42)

Simpson's Paradox

Player	1995 Hits/ At Bat	1995 BA	1996 Hits/ At Bat	1996 BA
Derek Jeter	12/48	0.25	183/582	0.314
David Justice	104/411	0.253	45/140	0.321

Higher Batting Average (BA)

Player	Comb Hits/ At Bat	Comb BA
Derek Jeter	195/630	0.31
David Justice	159/551	0.289

Higher Batting Average (BA)

Statistical Faux Pas (43)

Simpson's Paradox

Player	1995 Hits/ At Bat	1995 BA	1996 Hits/ At Bat	1996 BA	1997 Hits/ At Bat	1997 BA
Derek Jeter	12/48	0.25	183/582	0.314	190/654	0.291
David Justice	104/411	0.253	45/140	0.321	163/495	0.329

Higher Batting Average (BA)

Player	Comb Hits/ At Bat	Comb BA
Derek Jeter		
David Justice		

Statistical Faux Pas (44)

Simpson's Paradox

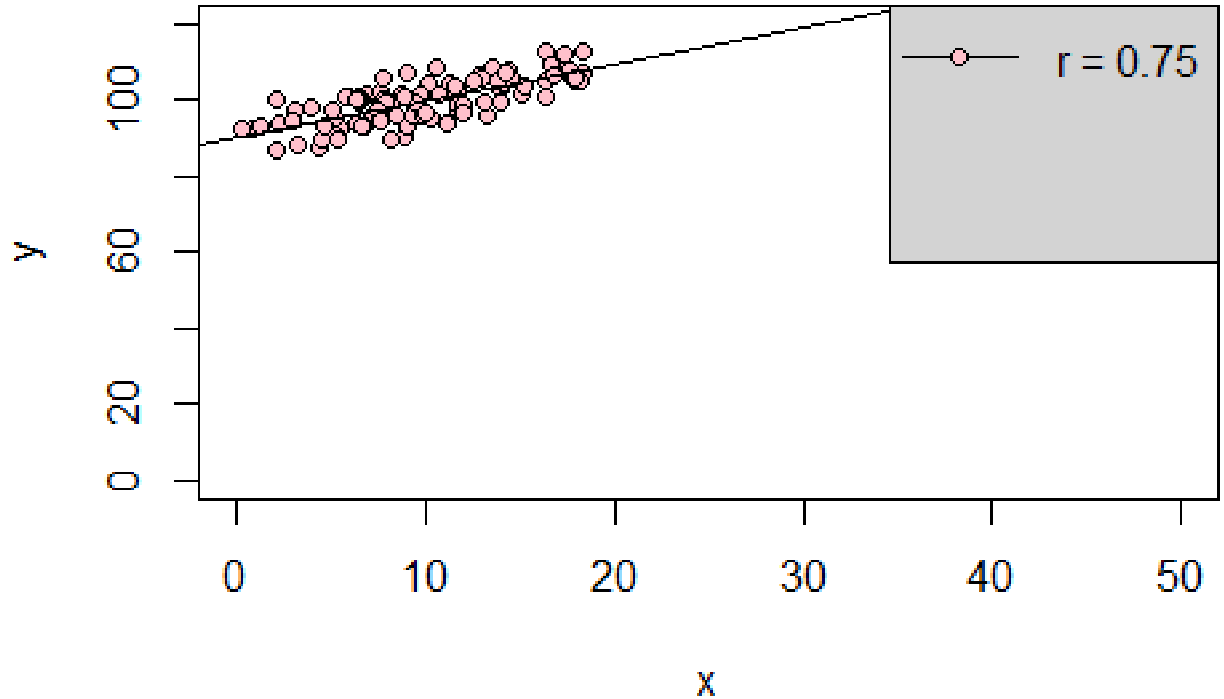
Player	1995 Hits/ At Bat	1995 BA	1996 Hits/ At Bat	1996 BA	1997 Hits/ At Bat	1997 BA
Derek Jeter	12/48	0.25	183/582	0.314	190/654	0.291
David Justice	104/411	0.253	45/140	0.321	163/495	0.329

Player	Comb Hits/ At Bat	Comb BA
Derek Jeter	385/1284	0.3
David Justice	312/1046	0.298

Higher Batting Average (BA)

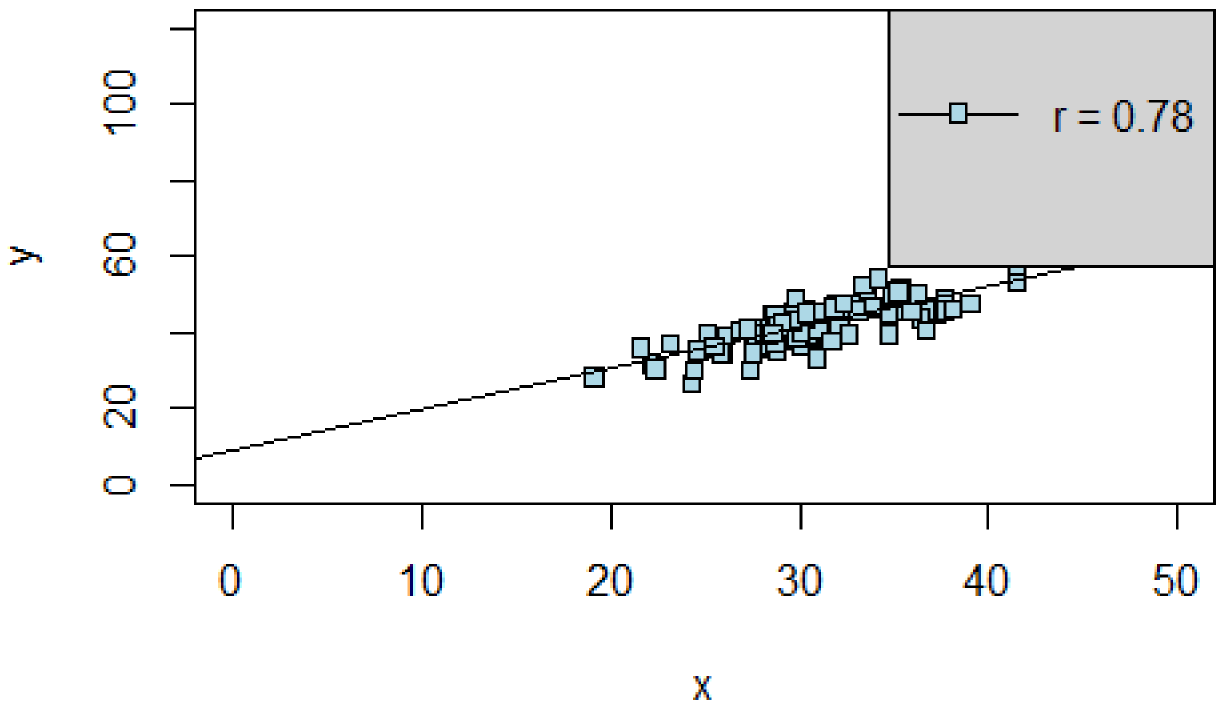
Statistical Faux Pas (45)

Simpson's Paradox



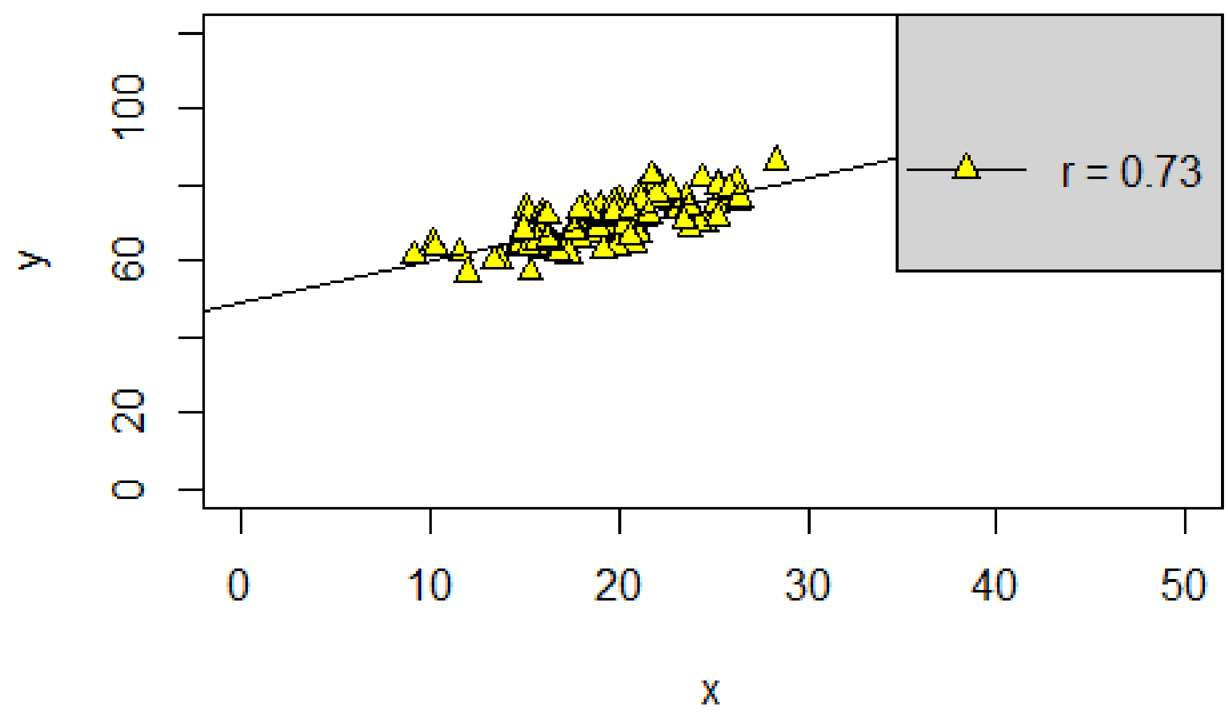
Statistical Faux Pas (46)

Simpson's Paradox



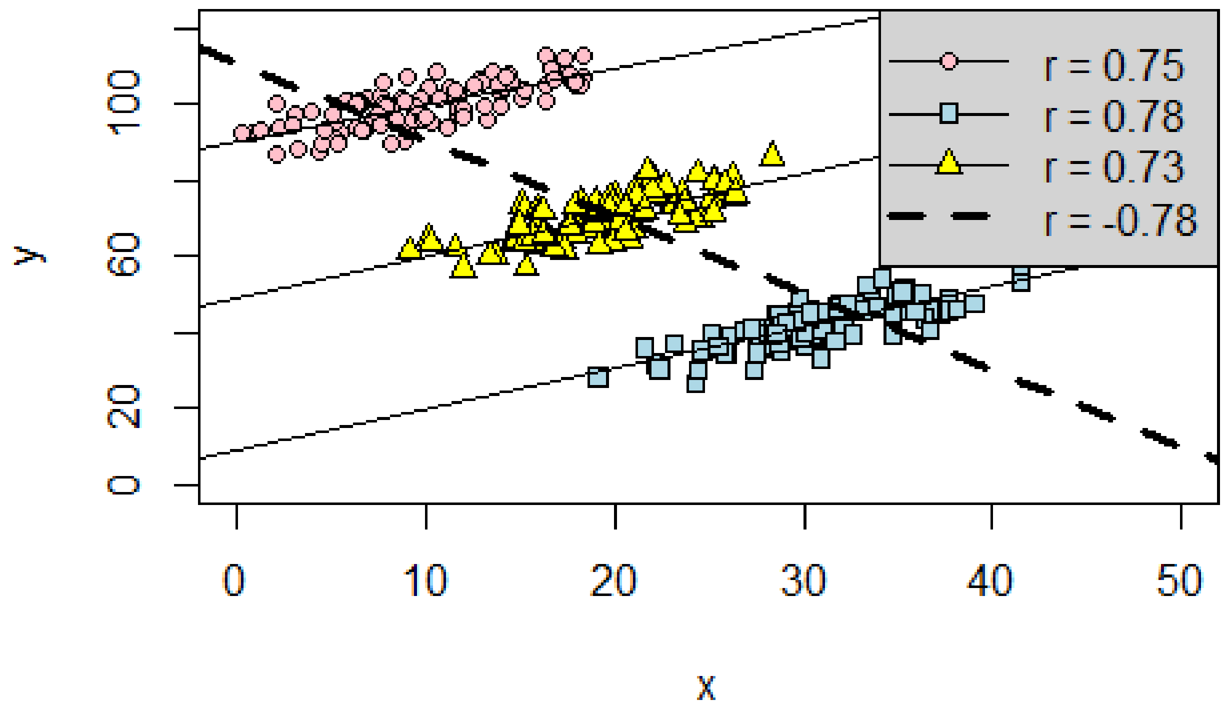
Statistical Faux Pas (47)

Simpson's Paradox



Statistical Faux Pas (48)

Simpson's Paradox



Statistical Faux Pas (49)

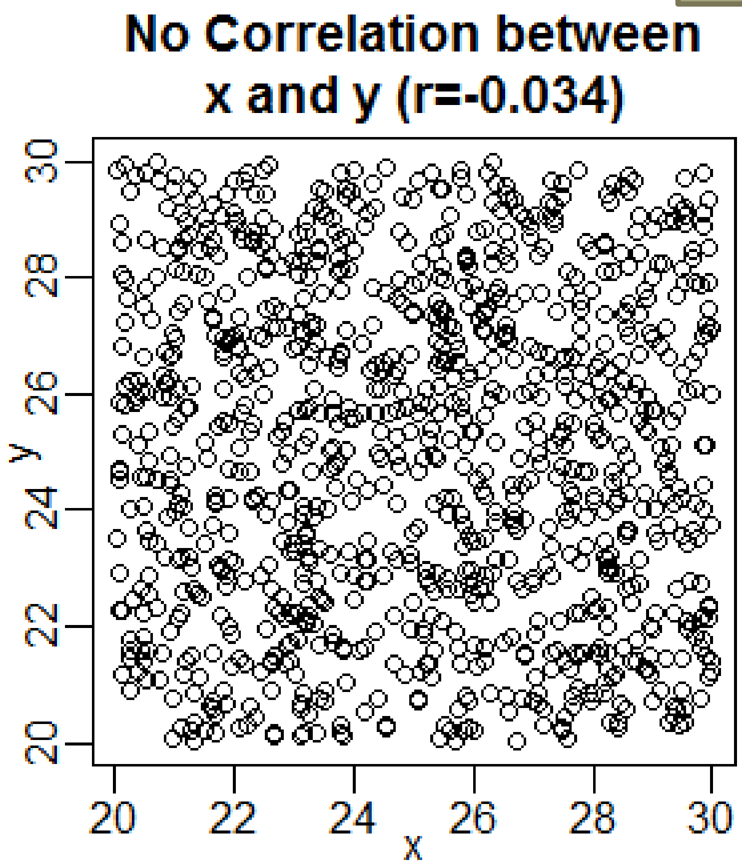
Spurious Correlation

- Correlation between ratios of absolute measurements
- Also Called:
  - Spurious Self-Correlation
  - Virtual Correlation
- https://en.wikipedia.org/wiki/Spurious_correlation
- See: SpuriousCorrelation.R

Statistical Faux Pas (50)

<u>Spurious Correlation</u>

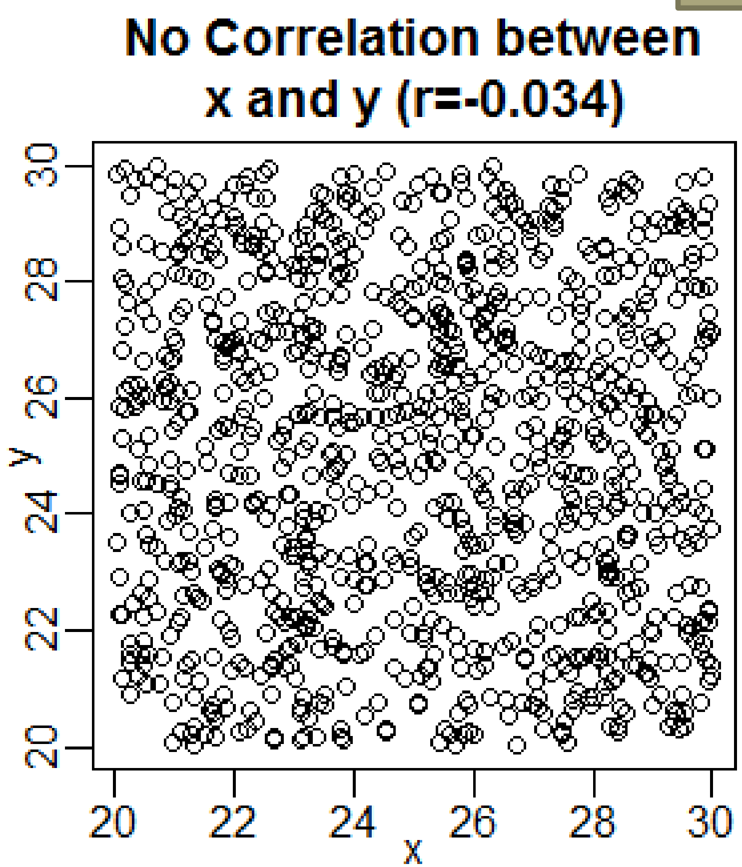
```
cor(x, y)      -0.034
```



Statistical Faux Pas (51)

Spurious Correlation

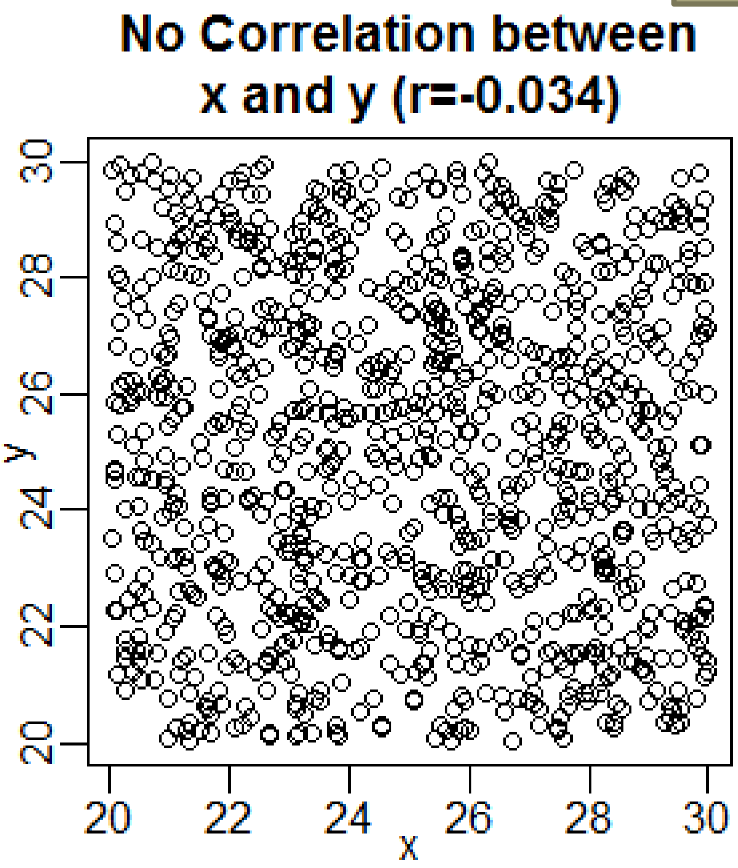
```
cor(x, y)     -0.034
cor(z, x)     -0.025
cor(z, y)      0.006
```

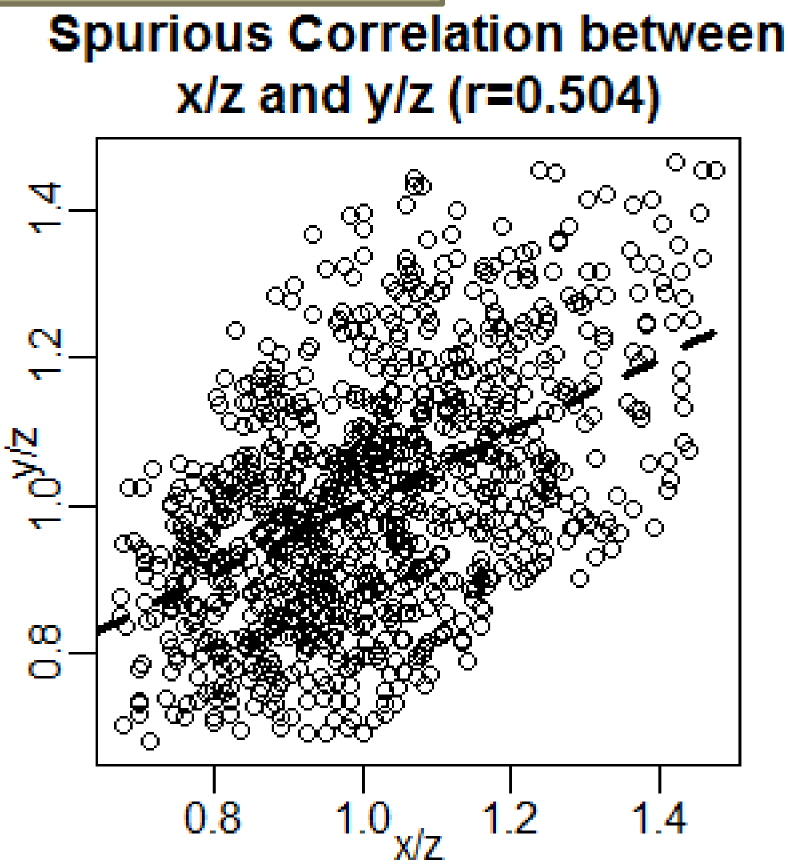


Statistical Faux Pas (52)

Spurious Correlation

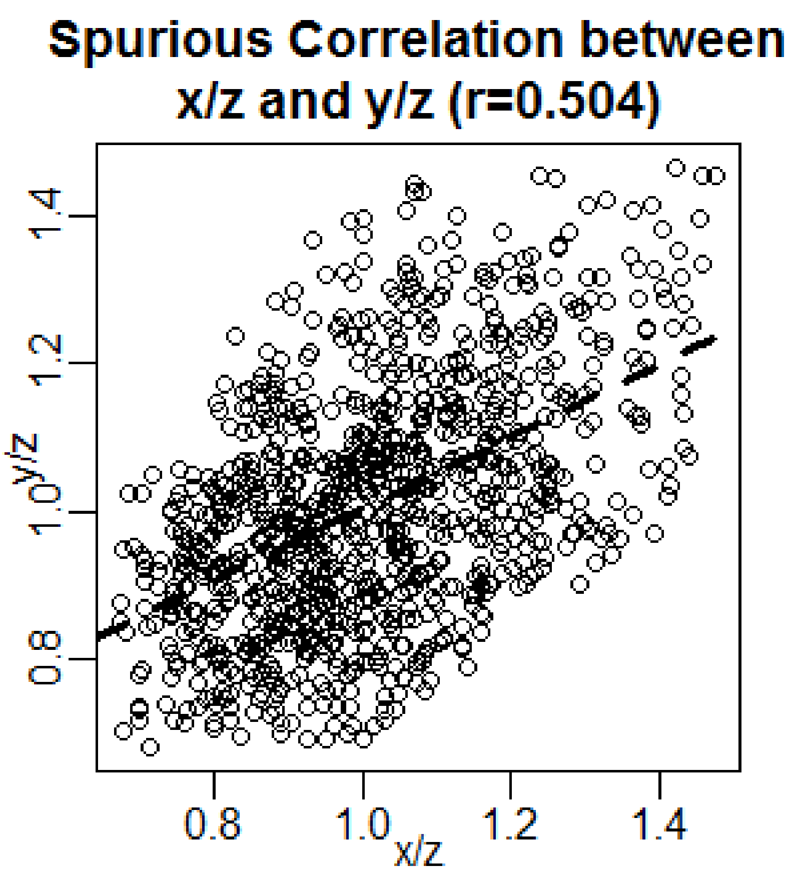
```
cor(x, y)       -0.034
cor(z, x)       -0.025
cor(z, y)        0.006
cor(x/z, y/z)  0.504
```





Statistical Faux Pas (53)

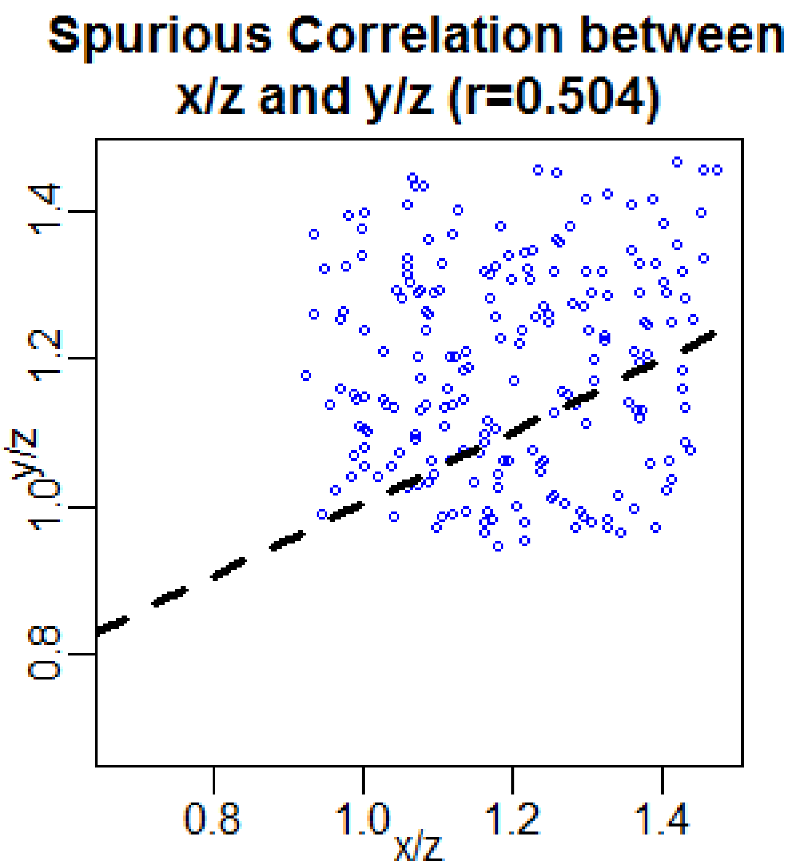
Spurious Correlation



Statistical Faux Pas (54)

Spurious Correlation

20 < z < 22
22 < z < 24
24 < z < 26
26 < z < 28
28 < z < 30



Statistical Faux Pas (55)

Spurious Correlation

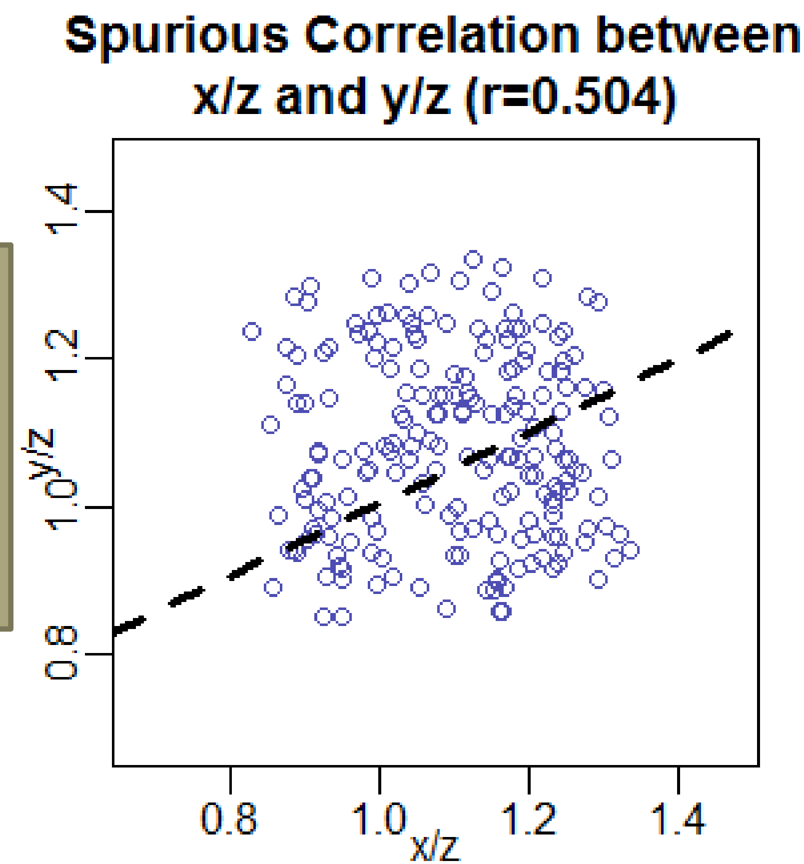
$20 < z < 22$

$22 < z < 24$

$24 < z < 26$

$26 < z < 28$

$28 < z < 30$



Statistical Faux Pas (56)

Spurious Correlation

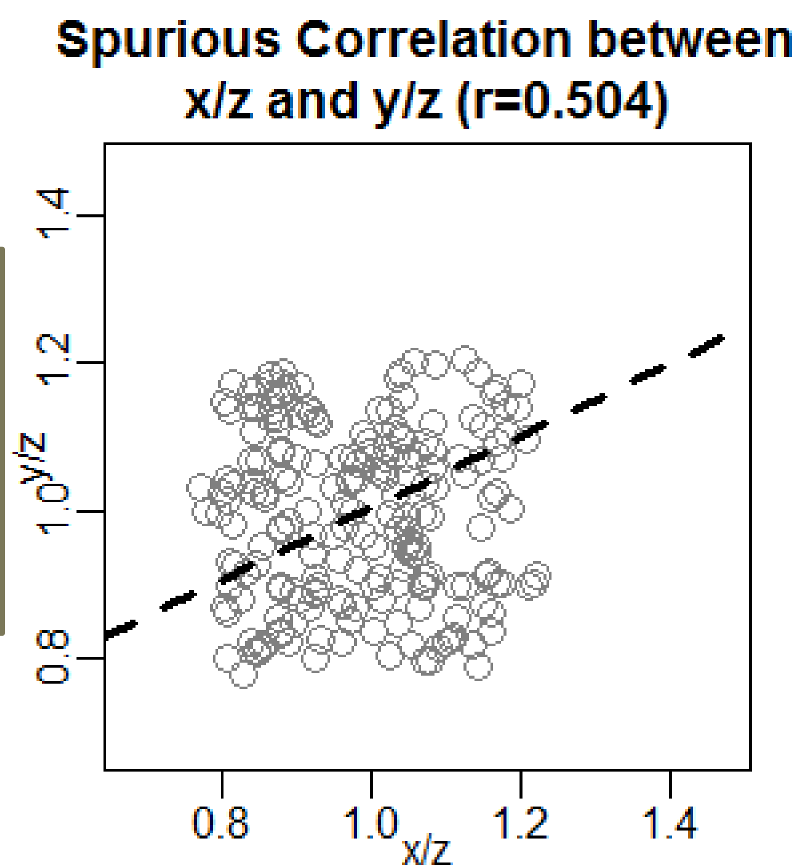
$20 < z < 22$

$22 < z < 24$

$24 < z < 26$

$26 < z < 28$

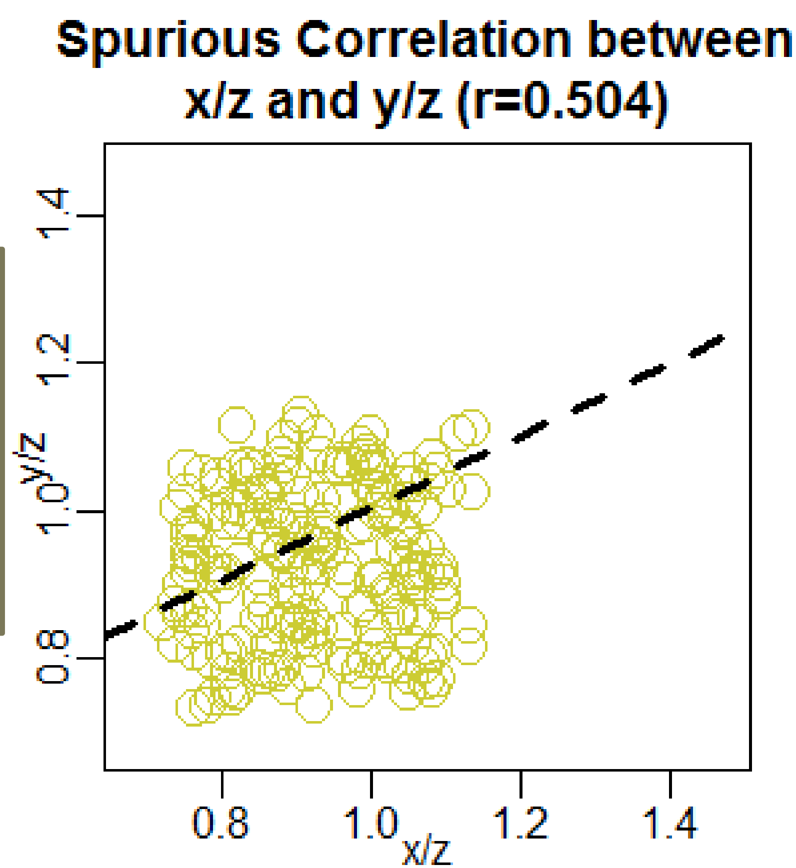
$28 < z < 30$



Statistical Faux Pas (57)

<u>Spurious Correlation</u>

20 < z < 22
22 < z < 24
24 < z < 26
26 < z < 28
28 < z < 30



Statistical Faux Pas (58)

Spurious Correlation

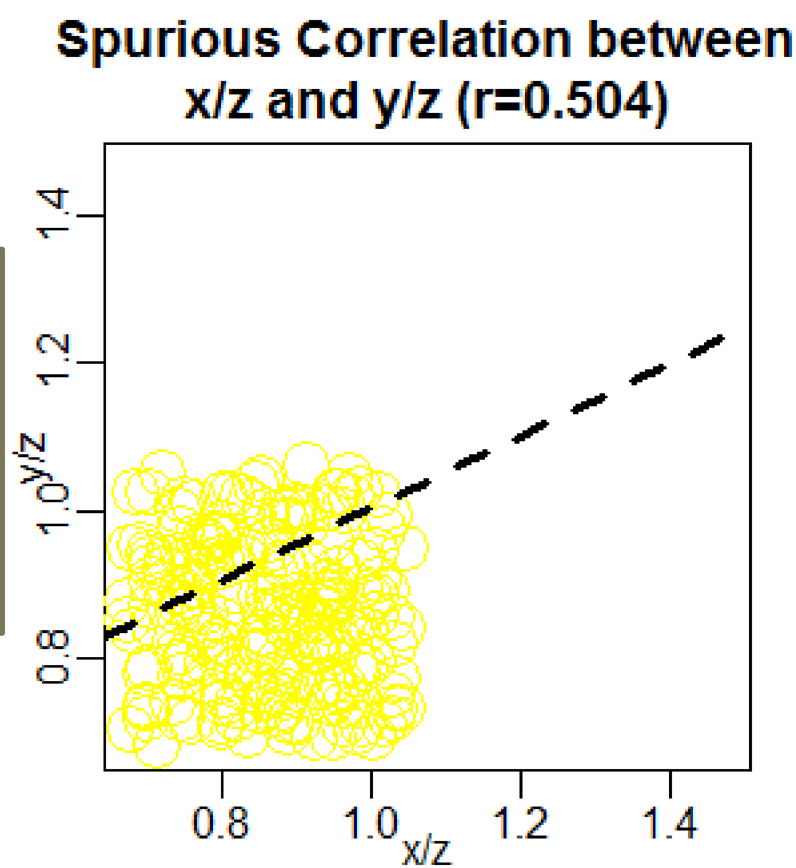
$20 < z < 22$

$22 < z < 24$

$24 < z < 26$

$26 < z < 28$

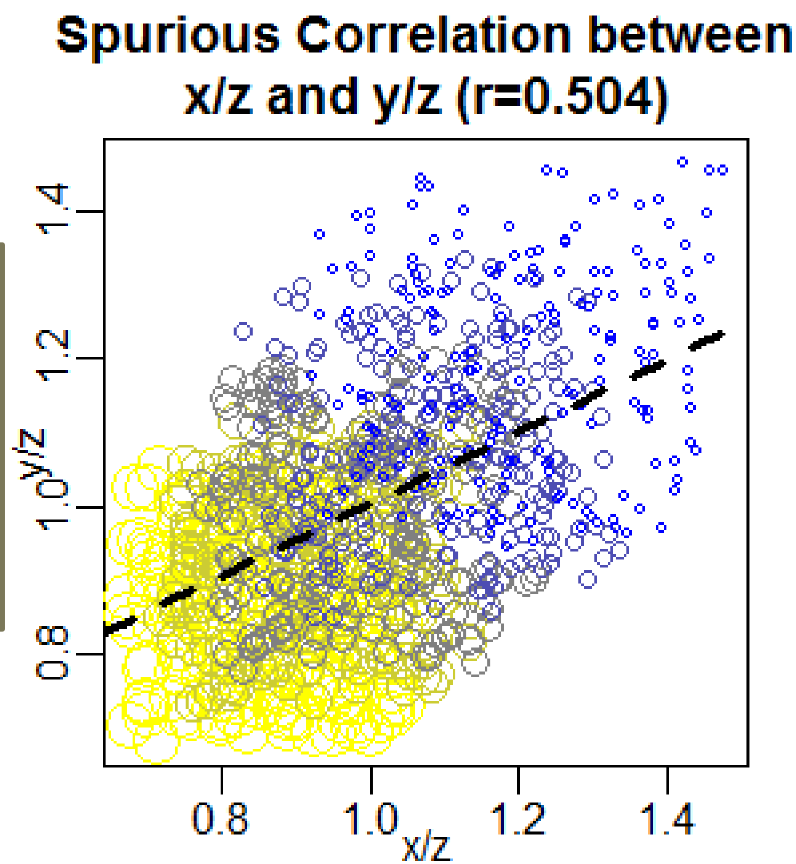
$28 < z < 30$



Statistical Faux Pas (59)

Spurious Correlation

20 < z < 22
22 < z < 24
24 < z < 26
26 < z < 28
28 < z < 30



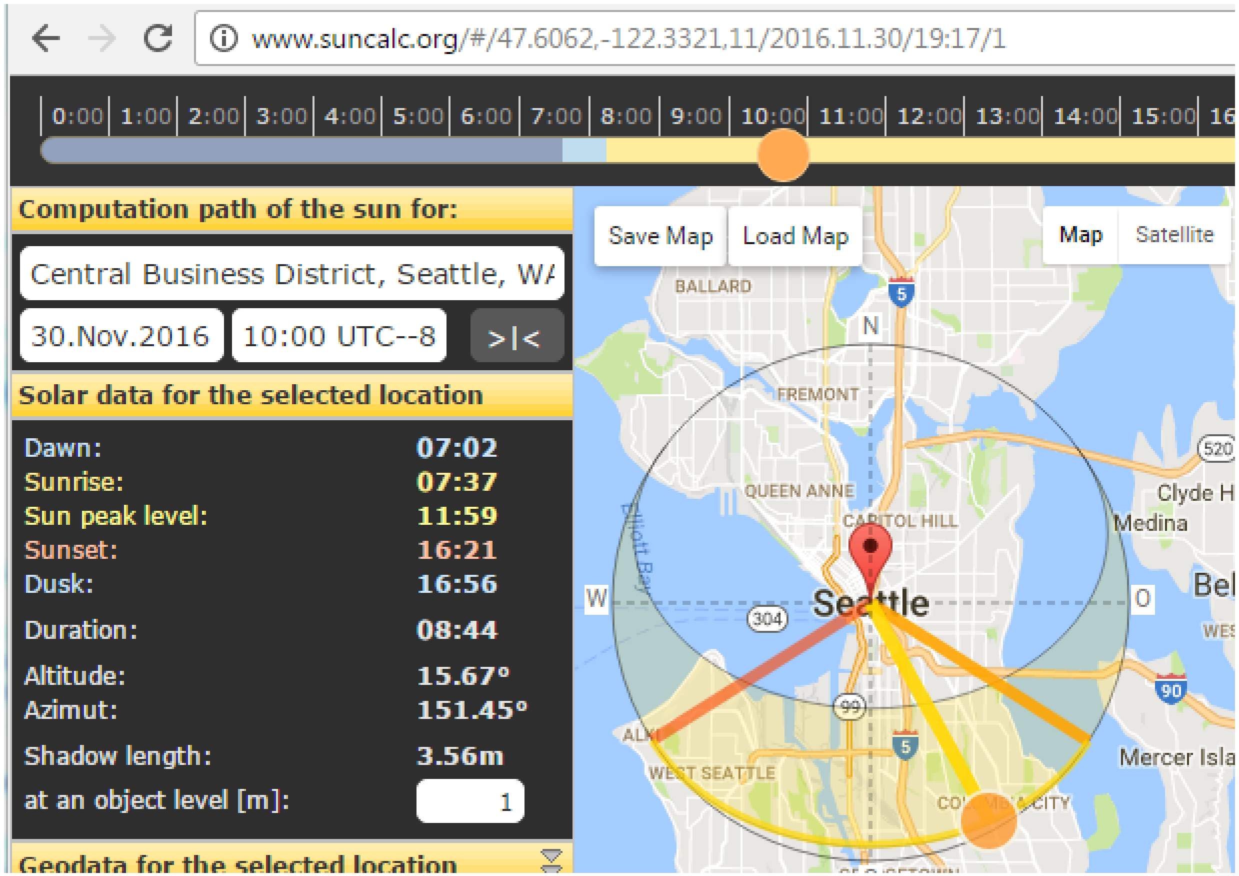
Statistical Faux Pas (60)

Example: Sun rise in Seattle (See: Sunrise.R)

- Given measurements between 7 AM and 1 PM, does the sun rise over time in Seattle?
- Height of sun is measured in degrees.
- We should get results from various times of year (i.e. summer solstice, fall equinox, winter solstice) and combine those data.

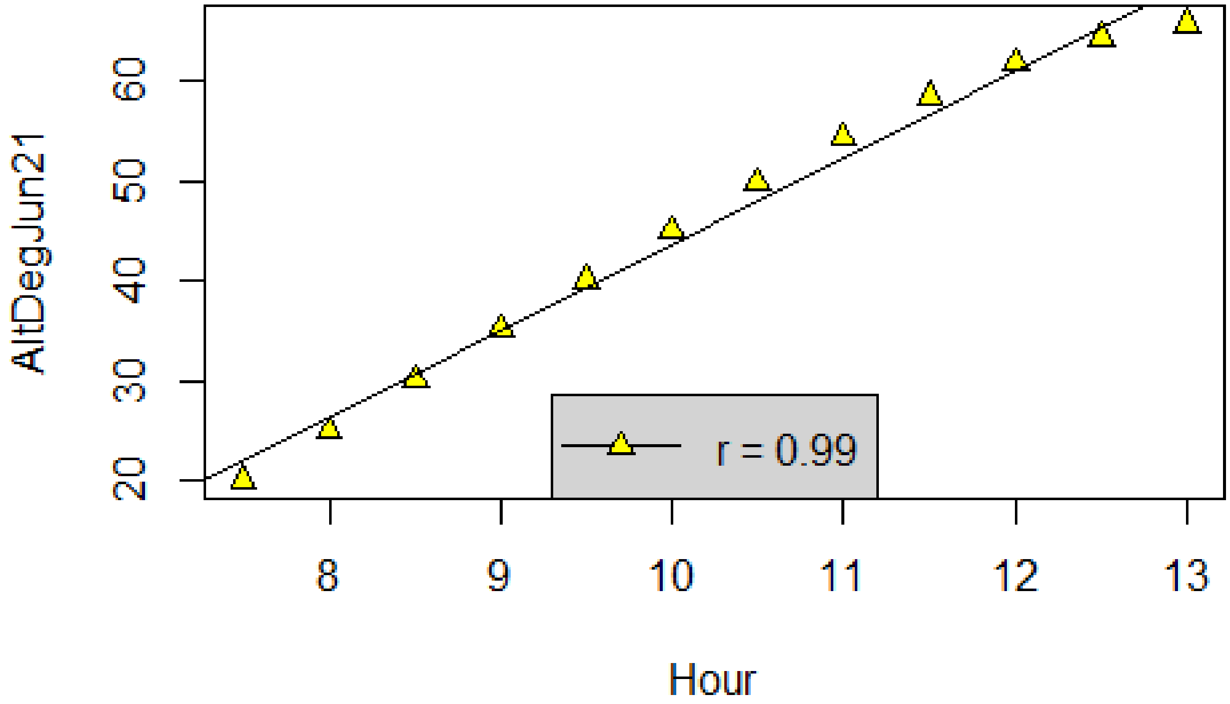
Statistical Faux Pas (61)

<u>Example: Sun rise in Seattle (See: Sunrise.R)</u>



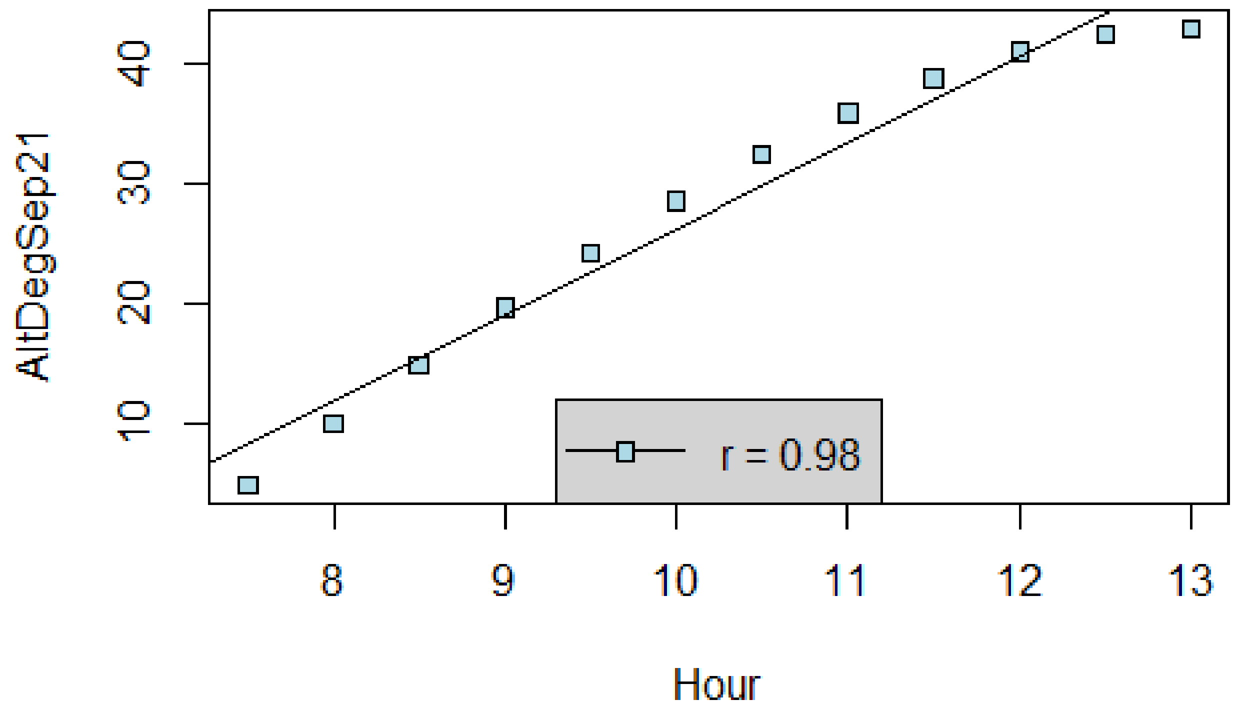
Statistical Faux Pas (62)

Example: Sun rise in Seattle (See: Sunrise.R)



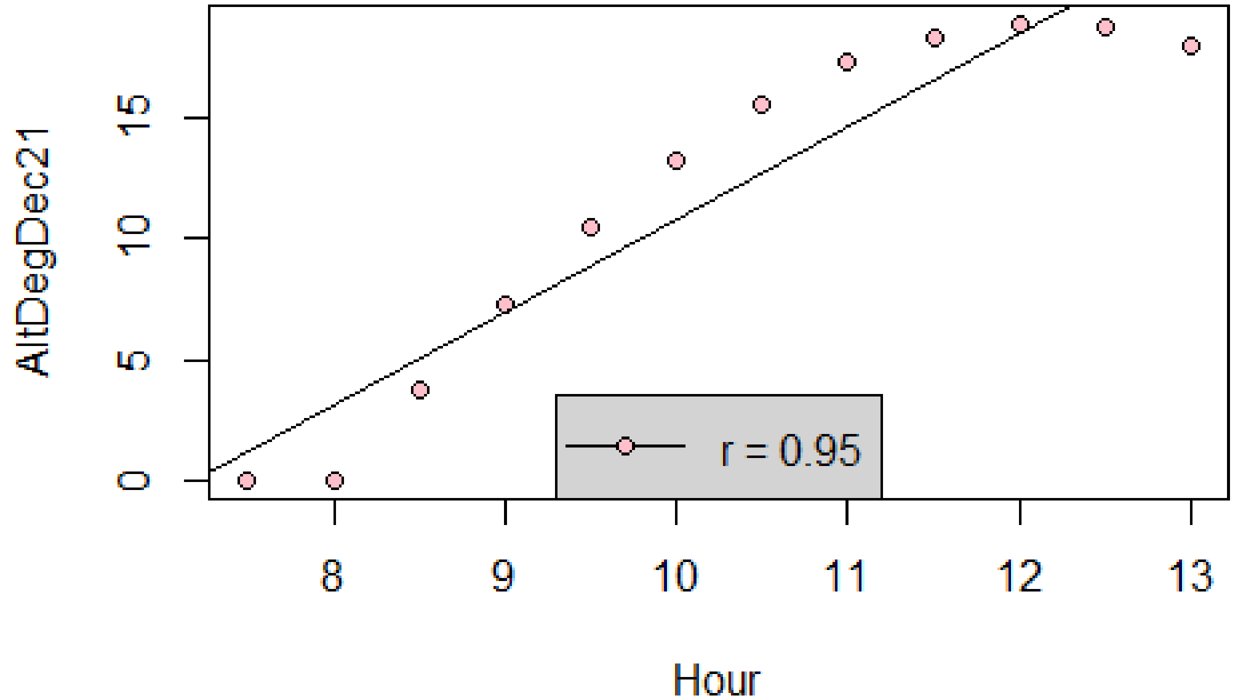
Statistical Faux Pas (63)

Example: Sun rise in Seattle (See: Sunrise.R)



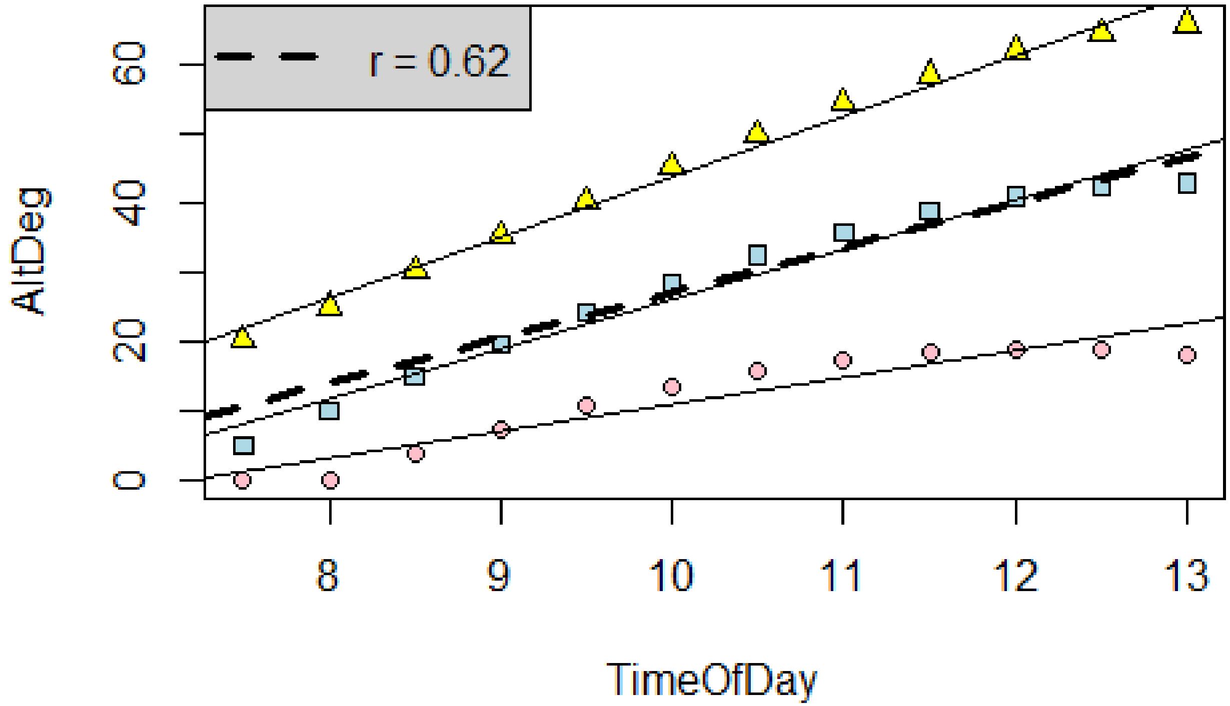
Statistical Faux Pas (64)

Example: Sun rise in Seattle (See: Sunrise.R)



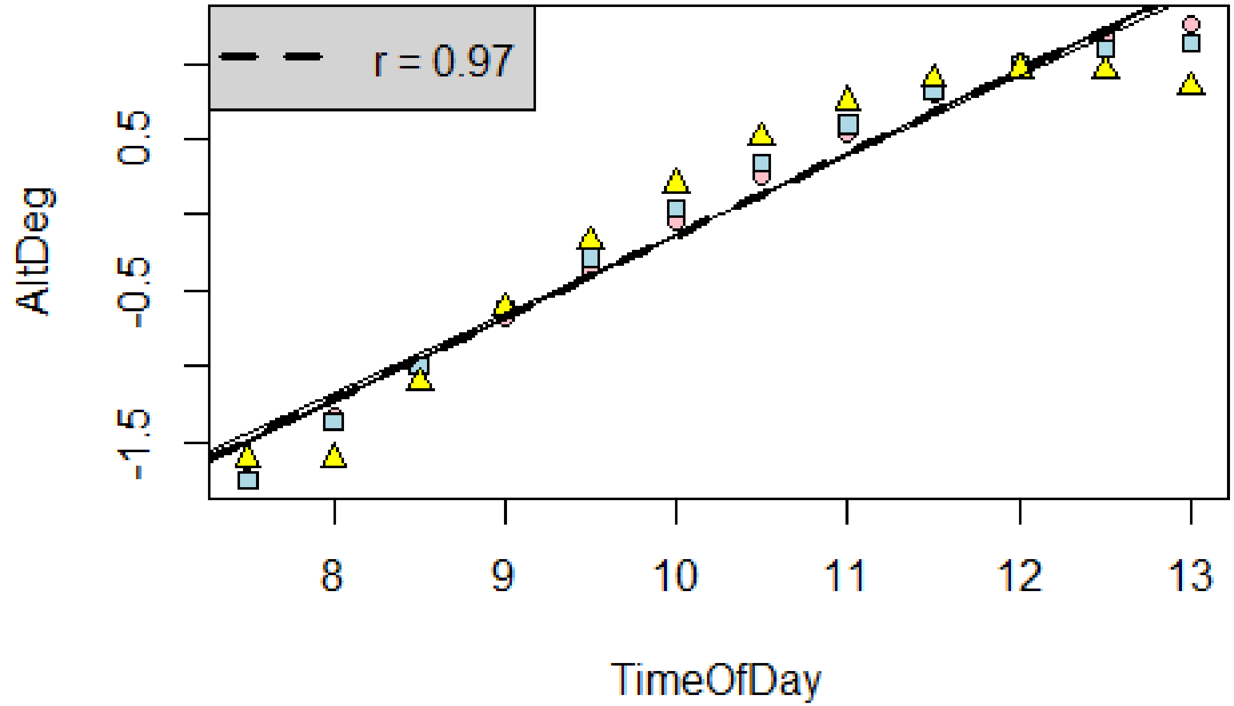
Statistical Faux Pas (65)

Example: Sun rise in Seattle (See: Sunrise.R)



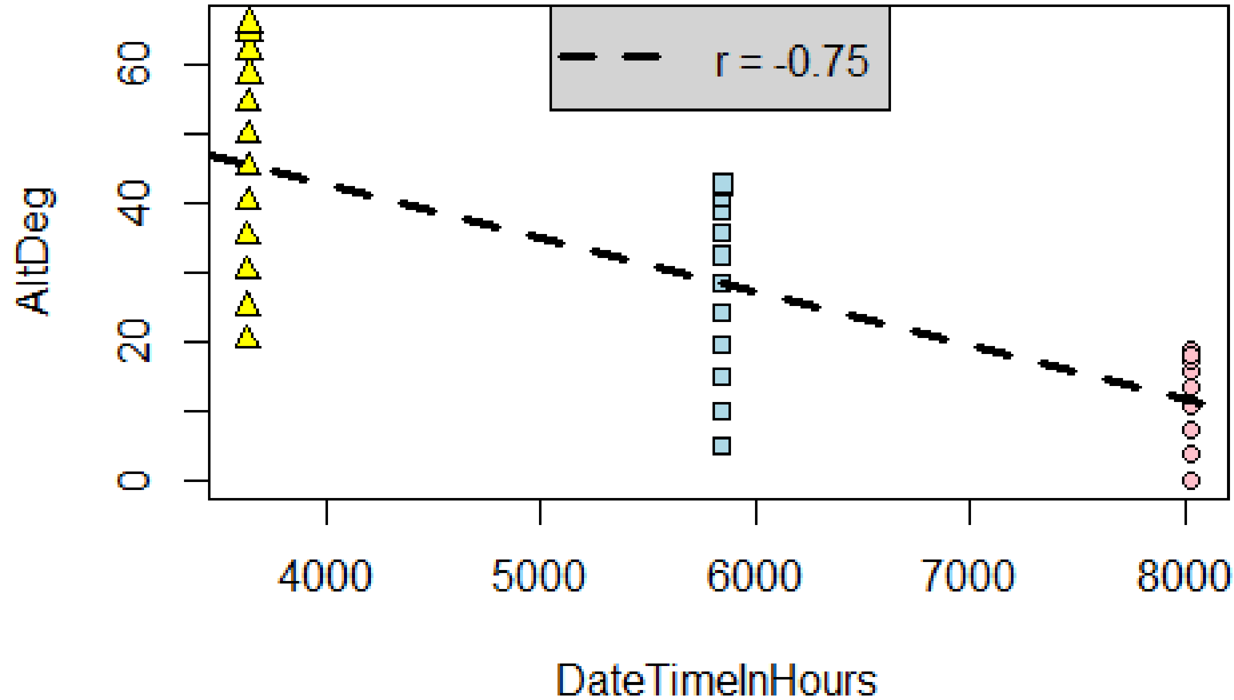
Statistical Faux Pas (66)

Example: Sun rise in Seattle (See: Sunrise.R)



Statistical Faux Pas (67)

Example: Sun rise in Seattle (See: Sunrise.R)



Statistical Faux Pas