Correlations

Recap

Population Variability

Sums of squares

$$SS = \Sigma(X_i - \mu_x)^2$$

Variance

$$\sigma^2 = \frac{\Sigma(X_i - \mu_x)^2}{N} = \frac{SS}{N}$$

Standard devation

$$\sigma = \sqrt{\frac{\Sigma(X_i - \mu_x)^2}{N}} = \sqrt{\frac{SS}{N}} = \sqrt{\sigma^2}$$

Sample variability

Sums of squares

$$SS = \Sigma(X_i - \bar{X})^2$$

Variance

$$s^2 = \frac{\Sigma(X_i - \bar{X})^2}{N - 1} = \frac{SS}{N - 1}$$

Standard devation

$$s = \sqrt{\frac{\Sigma(X_i - \bar{X})^2}{N - 1}} = \sqrt{\frac{SS}{N - 1}} = \sqrt{s^2}$$

Bi-variate descriptives

Covariation

"Sum of the cross-products"

Population

$$SP_{XY} = \Sigma(X_i - \mu_X)(Y_i - \mu_Y)$$

Sample

$$SP_{XY} = \Sigma(X_i - \bar{X})(Y_i - \bar{Y})$$

Covariance

Sort of like the variance of two variables

Population

$$\sigma_{XY} = \frac{\Sigma(X_i - \mu_X)(Y_i - \mu_Y)}{N}$$

Sample

$$s_{XY} = cov_{XY} = \frac{\Sigma(X_i - \bar{X})(Y_i - \bar{Y})}{N - 1}$$

Covariance table

$$\mathbf{K_{XX}} = \begin{bmatrix} \sigma^2_X & cov_{XY} & cov_{XZ} \\ cov_{YX} & \sigma^2_Y & cov_{YZ} \\ cov_{ZX} & cov_{ZY} & \sigma^2_Z \end{bmatrix}$$

Correlation

- Measure of association
- How much two variables are *linearly* related
- -1 to 1
- Sign indicates direction of relationship
- Invariant to changes in mean or scaling

Correlation

Pearson product moment correlation

Population

$$\rho_{XY} = \frac{\Sigma z_X z_Y}{N} = \frac{SP}{\sqrt{SS_X}\sqrt{SS_Y}} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$$

Sample

$$r_{XY} = \frac{\Sigma z_X z_Y}{n-1} = \frac{SP}{\sqrt{SS_X}\sqrt{SS_Y}} = \frac{s_{XY}}{s_X s_Y}$$

Conceptually

Ways to think about a correlation:

- How two vectors of numbers co-relate (i.e., parent & child heights)
- Product of z-scores (mathematically, it is)
- The average squared distance between 2 vectors in the same space

Effect size

- Recall that *z*-scores allow us to compare across units of measure; the products of standardized scores are themselves standardized.
- The correlation coefficient is a **standardized effect size** which can be used communicate the **strength** of a relationship.
- Correlations can be compared across studies, measures, constructs, time.
- Example: the correlation between age and height among children is $r = .70$.
- Building blocks of regression!

What is a large correlation?

- Cohen (1988): .1 (small), .3 (medium), .5 (large)
    - Often forgot: Cohen said only to use them when you had nothing else to go on, and has since regretted even suggesting benchmarks to begin with.
- Meyer & Hemphill said .3 is average

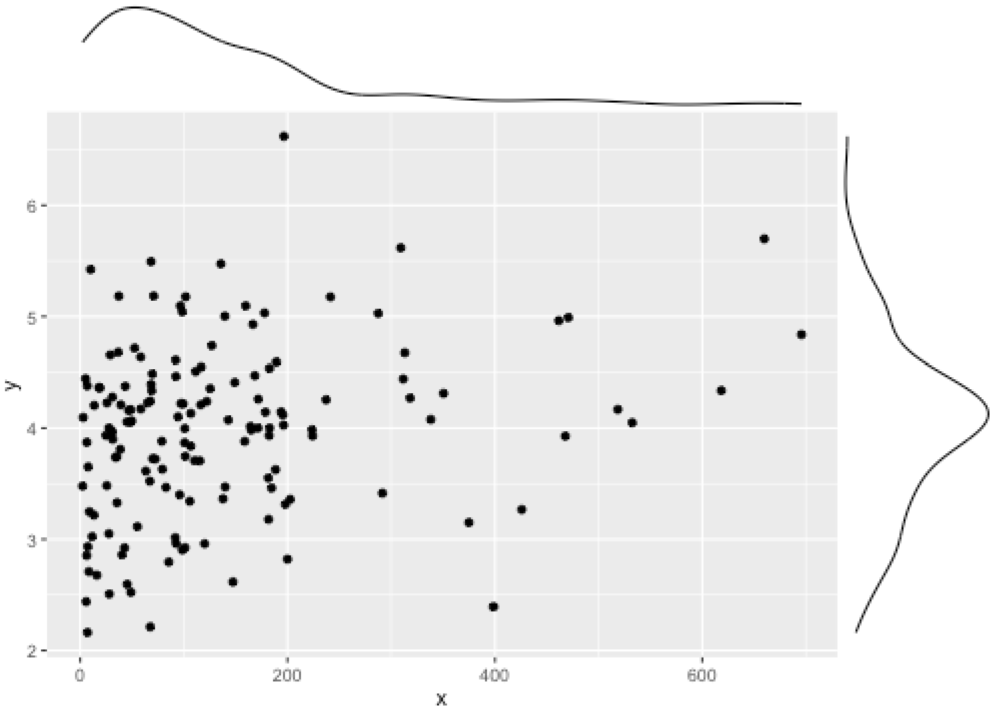
What affects correlations?

It's not enough to calculate a correlation between two variables. You should always look at a figure of the data to make sure the number accurately describes the relationship. Correlations can be easily fooled by qualities of your data, like:

1. Skewed distributions
2. Outliers
3. Restriction of range
4. Nonlinearity
5. Multiple Groups
6. Reliability

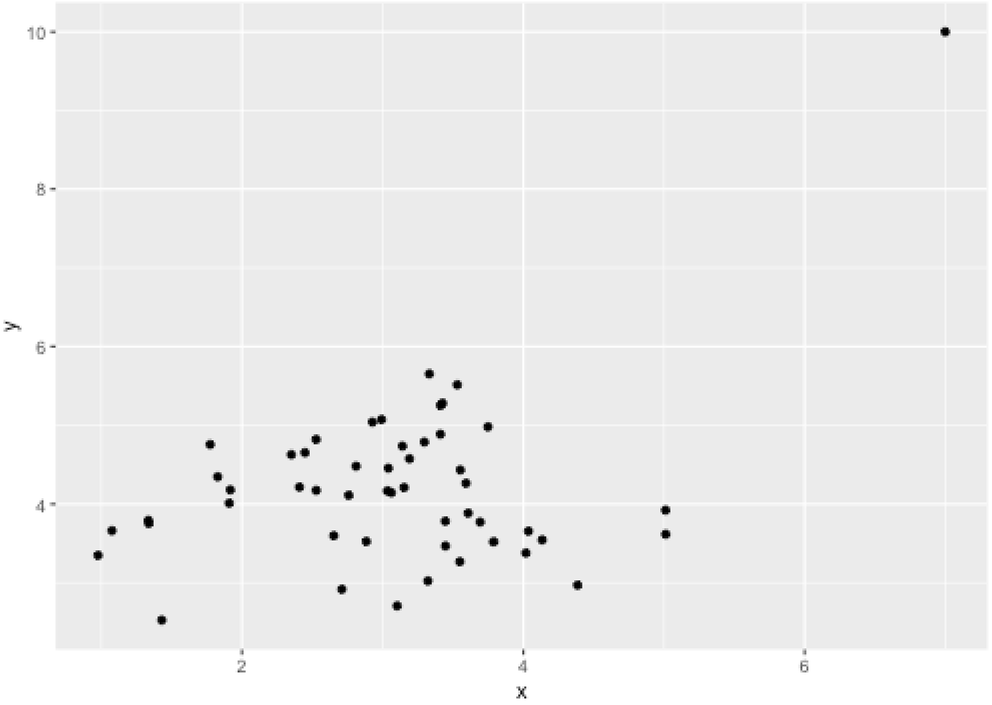
Skewed distributions

```
p = data %>% ggplot(aes(x=x, y=y)) + geom_point()
ggMarginal(p, type = "density")
```



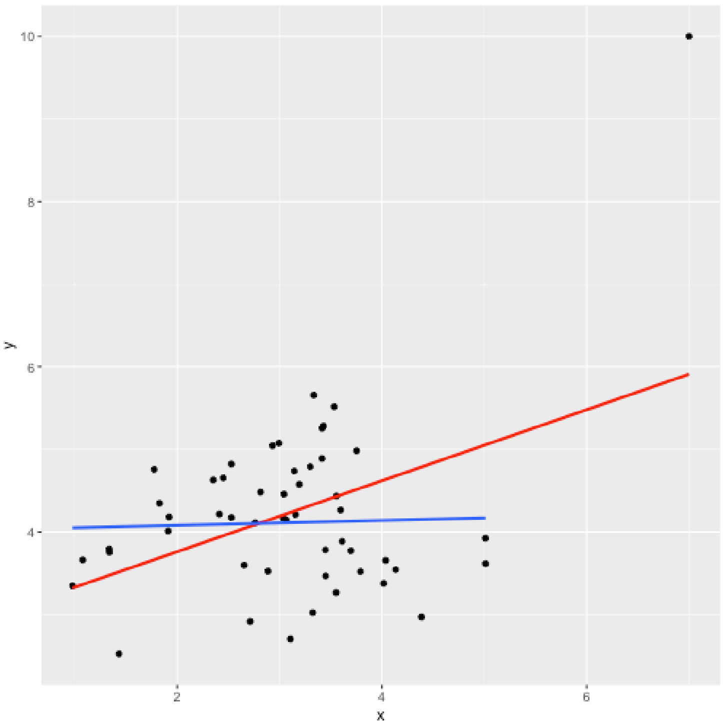
Outliers

```
data %>% ggplot(aes(x=x, y=y)) + geom_point()
```



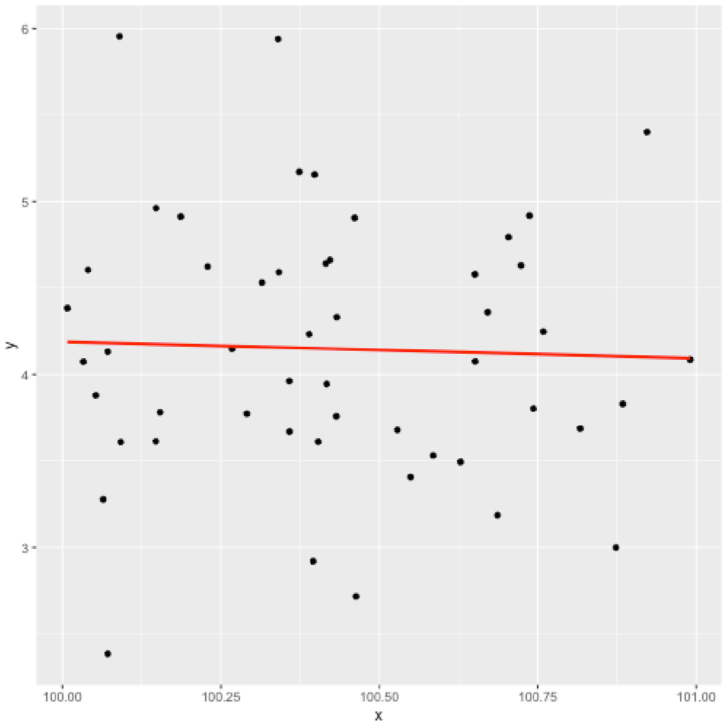
Outliers

```
data %>% ggplot(aes(x=x, y=y)) +
  geom_point() +
  geom_smooth(method = "lm",
              se = FALSE,
              color = "red") +
  geom_smooth(data = data[-51,],
              method = "lm",
              se = FALSE)
```



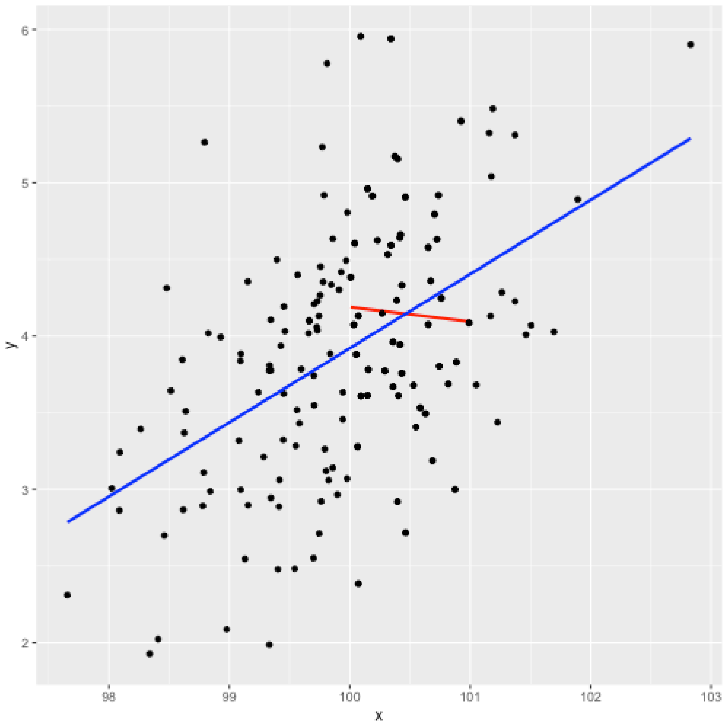
Restriction of range

```
data %>%
ggplot(aes(x=x, y=y)) +
  geom_point() +
  geom_smooth(method = "lm",
              se = FALSE,
              color = "red")
```



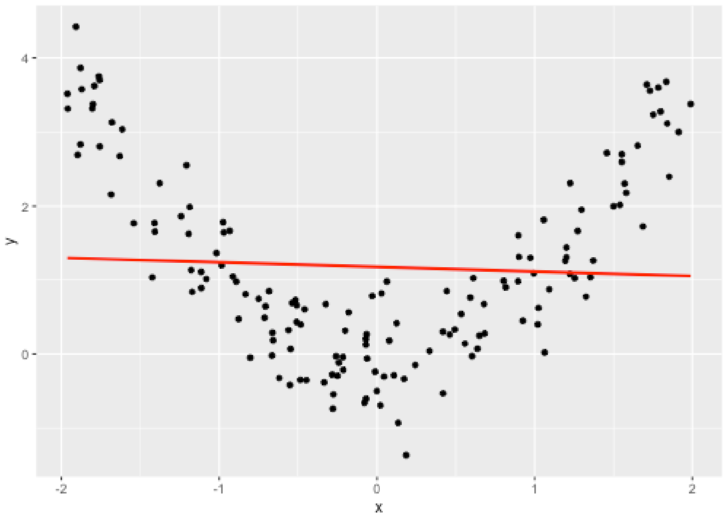
Restriction of range

```
data %>%
ggplot(aes(x=x, y=y)) +
  geom_point() +
  geom_smooth(method = "lm",
              se = FALSE,
              color = "red") +
  geom_point(data = real_data) +
  geom_smooth(method = "lm",
              se = FALSE,
              data = real_data,
              color = "blue")
```



Nonlinearity

```
data %>%
  ggplot(aes(x=x, y=y)) +
  geom_point() +
  geom_smooth(method = "lm",
              se = FALSE,
              color = "red")
```



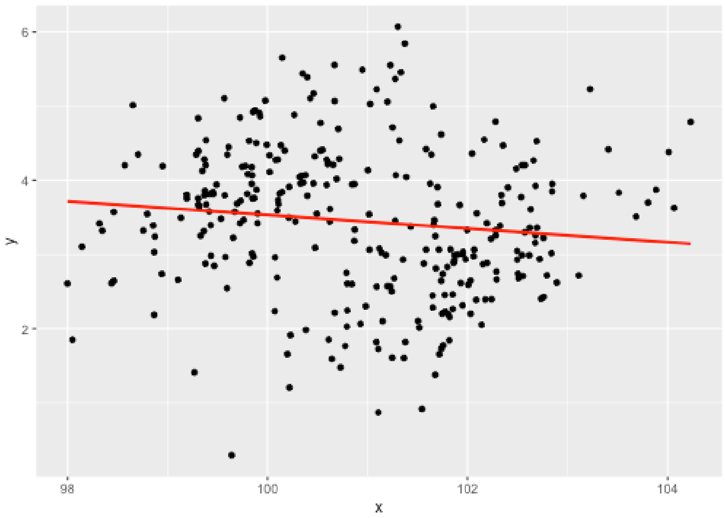
It's not always apparent

Sometimes issues that affect correlations won't appear in your graph, but you still need to know how to look for them.

- Multiple groups
- Low reliability (take a psychometrics or research methods course)

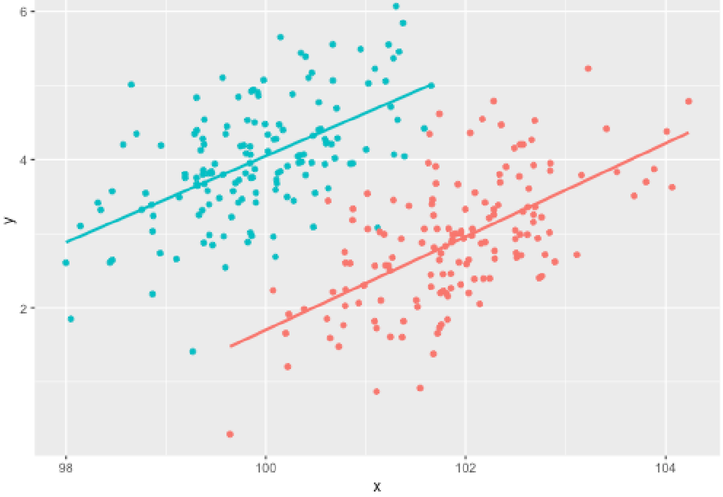
Multiple groups

```
data %>%
  ggplot(aes(x=x, y=y)) +
  geom_point() +
  geom_smooth(method = "lm", se =
```



Multiple groups

```
data %>%
  ggplot(aes(x=x, y=y,
             color = gender)) +
  geom_point() +
  geom_smooth(method = "lm", se =
  guides(color = F)
```



Reliability

Which would you rather have?

- 1-item final exam vs. a 30-item final exam?
- fMRI during a minor earthquake vs. no earthquake?
- Cognitive testing with Blue Angels flying vs. not with the Blue Angels?

Reliability

Which would you rather have?

- 1-item final exam vs. a 30-item final exam?
- fMRI during a minor earthquake vs. no earthquake?
- Cognitive testing with Blue Angels flying vs. not with the Blue Angels?

All measurement includes *error*

- Observed Score = True Score + Measurement Error

Reliability

- Error is random; it cannot correlate with anything
- Because we don't measure our variables perfectly, we get lower correlations compared to the "true" correlations
- Kind of analgous to power -- it's a ceiling
- If we want valid measures, they need to be reliable

So what do we do?

- If you're going to measure something, do it well
- Applies to *ALL* IVs and DVs, and all designs
- **Remember this when interpreting other research**

What is the size of the correlation?

- Chemotherapy and breast cancer survival?
- Batting ability and hit success on a single at bat?
- Antihistamine use and reduced sneezing/runny nose?
- Combat exposure and PTSD?
- Ibuprofen on pain reduction?
- Gender and weight?
- Therapy and well being?
- Observer ratings of attractiveness?
- Gender and arm strength?
- Nearness to equator and daily temperature for U.S.?

What is the size of the correlation?

- Chemotherapy and breast cancer survival? (.03)
- Batting ability and hit success on a single at bat? (.06)
- Antihistamine use and reduced sneezing/runny nose? (.11)
- Combat exposure and PTSD? (.11)
- Ibuprofen on pain reduction? (.14)
- Gender and weight? (.26)
- Therapy and well being? (.32)
- Observer ratings of attractiveness? (.39)
- Gender and arm strength? (.55)
- Nearness to equator and daily temperature for U.S.? (.60)

Questions to ask yourself:

- What is your N?
- What is the typical effect size in the field?
- Study design?
- What is your DV?
- Importance?

Correlation matrices

Correlations are both a descriptive and an inferential statistic. As a descriptive statistic, they're useful for understanding what's going on in a larger dataset.

Like we use the `summary()` or `describe()` (psych) functions to examine our dataset *before we run any infernetial tests*, we should also look at the correlation matrix.

```r
library(psych)
data(bfi)
head(bfi)
```

```
##       A1 A2 A3 A4 A5 C1 C2 C3 C4 C5 E1 E2 E3 E4 E5 N1 N2 N3 N4 N5 O1 O2 O3
## 61617  2  4  3  4  4  2  3  3  4  4  3  3  3  4  4  3  4  2  2  3  3  6  3
## 61618  2  4  5  2  5  5  4  4  3  4  1  1  6  4  3  3  3  3  5  5  4  2  4
## 61620  5  4  5  4  4  4  5  4  2  5  2  4  4  4  5  4  5  4  2  3  4  2  5
## 61621  4  4  6  5  5  4  4  3  5  5  5  3  4  4  4  2  5  2  4  1  3  3  4
## 61622  2  3  3  4  5  4  4  5  3  2  2  2  5  4  5  2  3  4  4  3  3  3  4
## 61623  6  6  5  6  5  6  6  6  1  3  2  1  6  5  6  3  5  2  2  3  4  3  5
##       O5 gender education age
## 61617  3      1        NA  16
## 61618  3      2        NA  18
## 61620  2      2        NA  17
## 61621  5      2        NA  17
## 61622  3      1        NA  17
## 61623  1      2         3  21
```

```
cor(bfi)
```

```
##             A1 A2 A3 A4 A5 C1 C2 C3 C4 C5 E1 E2 E3 E4 E5 N1 N2 N3 N4 N5 O1
## A1           1 NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA
## A2          NA  1 NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA
## A3          NA NA  1 NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA
## A4          NA NA NA  1 NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA
## A5          NA NA NA NA  1 NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA
## C1          NA NA NA NA NA  1 NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA
## C2          NA NA NA NA NA NA  1 NA NA NA NA NA NA NA NA NA NA NA NA NA NA
## C3          NA NA NA NA NA NA NA  1 NA NA NA NA NA NA NA NA NA NA NA NA NA
## C4          NA NA NA NA NA NA NA NA  1 NA NA NA NA NA NA NA NA NA NA NA NA
## C5          NA NA NA NA NA NA NA NA NA  1 NA NA NA NA NA NA NA NA NA NA NA
## E1          NA NA NA NA NA NA NA NA NA NA  1 NA NA NA NA NA NA NA NA NA NA
## E2          NA NA NA NA NA NA NA NA NA NA NA  1 NA NA NA NA NA NA NA NA NA
## E3          NA NA NA NA NA NA NA NA NA NA NA NA  1 NA NA NA NA NA NA NA NA
## E4          NA NA NA NA NA NA NA NA NA NA NA NA NA  1 NA NA NA NA NA NA NA
## E5          NA NA NA NA NA NA NA NA NA NA NA NA NA NA  1 NA NA NA NA NA NA
## N1          NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA  1 NA NA NA NA NA
## N2          NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA  1 NA NA NA NA
## N3          NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA  1 NA NA NA
## N4          NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA  1 NA NA
## N5          NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA  1 NA
## O1          NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA  1
## O2          NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA
## O3          NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA
## O4          NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA
## O5          NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA
```

```r
round(cor(bfi, use = "pairwise"),2)
```

```
##             A1    A2    A3    A4    A5    C1    C2    C3    C4    C5    E1
## A1        1.00 -0.34 -0.27 -0.15 -0.18  0.03  0.02 -0.02  0.13  0.05  0.11
## A2       -0.34  1.00  0.49  0.34  0.39  0.09  0.14  0.19 -0.15 -0.12 -0.21
## A3       -0.27  0.49  1.00  0.36  0.50  0.10  0.14  0.13 -0.12 -0.16 -0.21
## A4       -0.15  0.34  0.36  1.00  0.31  0.09  0.23  0.13 -0.15 -0.24 -0.11
## A5       -0.18  0.39  0.50  0.31  1.00  0.12  0.11  0.13 -0.13 -0.17 -0.25
## C1        0.03  0.09  0.10  0.09  0.12  1.00  0.43  0.31 -0.34 -0.25 -0.02
## C2        0.02  0.14  0.14  0.23  0.11  0.43  1.00  0.36 -0.38 -0.30  0.02
## C3       -0.02  0.19  0.13  0.13  0.13  0.31  0.36  1.00 -0.34 -0.34  0.00
## C4        0.13 -0.15 -0.12 -0.15 -0.13 -0.34 -0.38 -0.34  1.00  0.48  0.09
## C5        0.05 -0.12 -0.16 -0.24 -0.17 -0.25 -0.30 -0.34  0.48  1.00  0.06
## E1        0.11 -0.21 -0.21 -0.11 -0.25 -0.02  0.02  0.00  0.09  0.06  1.00
## E2        0.09 -0.23 -0.29 -0.19 -0.33 -0.09 -0.06 -0.08  0.20  0.26  0.47
## E3       -0.05  0.25  0.39  0.19  0.42  0.12  0.15  0.09 -0.08 -0.16 -0.33
## E4       -0.06  0.28  0.38  0.30  0.47  0.14  0.12  0.09 -0.11 -0.20 -0.42
## E5       -0.02  0.29  0.25  0.16  0.27  0.25  0.25  0.21 -0.24 -0.23 -0.30
## N1        0.17 -0.09 -0.08 -0.10 -0.20 -0.07 -0.02 -0.07  0.22  0.21  0.02
## N2        0.14 -0.05 -0.09 -0.14 -0.19 -0.04 -0.01 -0.06  0.16  0.25  0.01
## N3        0.10 -0.04 -0.04 -0.07 -0.14 -0.03  0.00 -0.07  0.21  0.24  0.05
## N4        0.05 -0.09 -0.13 -0.17 -0.20 -0.10 -0.05 -0.11  0.26  0.34  0.23
## N5        0.02  0.02 -0.04 -0.01 -0.08 -0.05  0.05 -0.01  0.20  0.17  0.05
## O1        0.01  0.13  0.15  0.06  0.16  0.17  0.16  0.09 -0.09 -0.08 -0.10
## O2        0.08  0.02  0.00  0.04  0.00 -0.11 -0.04 -0.03  0.21  0.14  0.04
## O3       -0.06  0.16  0.22  0.07  0.24  0.19  0.19  0.06 -0.08 -0.08 -0.22
## O4       -0.08  0.09  0.04 -0.04  0.02  0.11  0.06  0.02  0.05  0.14  0.08
```

```r
round(cor(bfi, use = "complete"),2)
```

```
##                A1    A2    A3    A4    A5    C1    C2    C3    C4    C5    E1
## A1           1.00 -0.34 -0.26 -0.14 -0.19  0.02  0.01 -0.01  0.10  0.02  0.12
## A2          -0.34  1.00  0.48  0.34  0.38  0.09  0.13  0.19 -0.14 -0.11 -0.24
## A3          -0.26  0.48  1.00  0.38  0.50  0.10  0.14  0.13 -0.12 -0.15 -0.22
## A4          -0.14  0.34  0.38  1.00  0.32  0.08  0.22  0.13 -0.16 -0.24 -0.14
## A5          -0.19  0.38  0.50  0.32  1.00  0.12  0.11  0.13 -0.12 -0.16 -0.25
## C1           0.02  0.09  0.10  0.08  0.12  1.00  0.43  0.32 -0.35 -0.25 -0.03
## C2           0.01  0.13  0.14  0.22  0.11  0.43  1.00  0.36 -0.38 -0.30  0.02
## C3          -0.01  0.19  0.13  0.13  0.13  0.32  0.36  1.00 -0.35 -0.35 -0.02
## C4           0.10 -0.14 -0.12 -0.16 -0.12 -0.35 -0.38 -0.35  1.00  0.48  0.10
## C5           0.02 -0.11 -0.15 -0.24 -0.16 -0.25 -0.30 -0.35  0.48  1.00  0.07
## E1           0.12 -0.24 -0.22 -0.14 -0.25 -0.03  0.02 -0.02  0.10  0.07  1.00
## E2           0.08 -0.24 -0.29 -0.20 -0.33 -0.10 -0.07 -0.09  0.21  0.26  0.47
## E3          -0.04  0.25  0.38  0.20  0.41  0.13  0.15  0.10 -0.09 -0.17 -0.33
## E4          -0.07  0.30  0.39  0.33  0.48  0.14  0.12  0.10 -0.12 -0.21 -0.42
## E5          -0.02  0.30  0.26  0.16  0.27  0.26  0.25  0.22 -0.23 -0.24 -0.31
## N1           0.16 -0.08 -0.07 -0.09 -0.19 -0.06 -0.02 -0.08  0.21  0.21  0.01
## N2           0.13 -0.04 -0.08 -0.15 -0.19 -0.03  0.00 -0.06  0.15  0.24  0.01
## N3           0.09 -0.02 -0.03 -0.07 -0.13 -0.01  0.01 -0.07  0.20  0.23  0.05
## N4           0.04 -0.09 -0.13 -0.16 -0.21 -0.09 -0.04 -0.13  0.28  0.35  0.23
## N5           0.01  0.02 -0.04  0.00 -0.08 -0.05  0.05 -0.04  0.21  0.18  0.04
## O1           0.00  0.11  0.14  0.04  0.15  0.18  0.16  0.09 -0.10 -0.09 -0.10
## O2           0.07  0.03  0.03  0.05  0.00 -0.13 -0.05 -0.03  0.21  0.12  0.06
## O3          -0.06  0.15  0.22  0.04  0.22  0.19  0.18  0.06 -0.07 -0.07 -0.21
## O4          -0.09  0.05  0.02 -0.06  0.00  0.08  0.03  0.00  0.07  0.14  0.08
```

Pairwise vs. Listwise Deletion

With **pairwise deletion**, different sets of cases contribute to different correlations. That maximizes the sample sizes, but can lead to problems if the data are missing for some systematic reason.

Listwise deletion (often referred to in R as use complete cases) doesn't have the same issue of biasing correlations, but does result in smaller samples and potentially limited generalizability.

A good practice is comparing the different matrices; if the correlation values are very different, this suggests that the missingness that affects pairwise deletion is systematic.

```r
round(cor(bfi, use = "pairwise")- cor(bfi, use = "complete"),2)
```

```
##               A1    A2    A3    A4    A5    C1    C2    C3    C4    C5    E1
## A1          0.00  0.00  0.00  0.00  0.00  0.01  0.00 -0.01  0.03  0.03 -0.01
## A2          0.00  0.00  0.00 -0.01  0.01  0.00  0.01  0.01 -0.01 -0.01  0.03
## A3          0.00  0.00  0.00 -0.02  0.00  0.00  0.00  0.00  0.00 -0.01  0.00
## A4          0.00 -0.01 -0.02  0.00 -0.01  0.01  0.01  0.00  0.01  0.00  0.03
## A5          0.00  0.01  0.00 -0.01  0.00  0.00  0.00  0.00 -0.01 -0.01  0.00
## C1          0.01  0.00  0.00  0.01  0.00  0.00  0.00 -0.01  0.01  0.00  0.00
## C2          0.00  0.01  0.00  0.01  0.00  0.00  0.00  0.00  0.00  0.00 -0.01
## C3         -0.01  0.01  0.00  0.00  0.00 -0.01  0.00  0.00  0.02  0.01  0.02
## C4          0.03 -0.01  0.00  0.01 -0.01  0.01  0.00  0.02  0.00 -0.01 -0.01
## C5          0.03 -0.01 -0.01  0.00 -0.01  0.00  0.00  0.01 -0.01  0.00  0.00
## E1         -0.01  0.03  0.00  0.03  0.00  0.00 -0.01  0.02 -0.01  0.00  0.00
## E2          0.01  0.01  0.00  0.01  0.00  0.01  0.01  0.01 -0.01  0.00  0.00
## E3          0.00  0.00  0.00 -0.01  0.00 -0.02  0.00 -0.02  0.01  0.01  0.01
## E4          0.01 -0.02 -0.02 -0.03 -0.01  0.00  0.00 -0.01  0.01  0.01  0.00
## E5          0.00  0.00 -0.01  0.00  0.00 -0.01  0.00  0.00  0.00  0.01  0.00
## N1          0.01 -0.01 -0.02  0.00  0.00 -0.01  0.00  0.01  0.01  0.01  0.01
## N2          0.01 -0.01  0.00  0.00  0.00 -0.01 -0.01  0.00  0.01  0.01  0.01
## N3          0.01 -0.02 -0.01  0.00 -0.01 -0.02 -0.01  0.01  0.01  0.01  0.00
## N4          0.01  0.00  0.00 -0.01  0.01 -0.01 -0.01  0.02 -0.02 -0.01  0.00
## N5          0.01  0.00  0.00  0.00  0.00  0.00  0.00  0.02 -0.02 -0.01  0.01
## O1          0.01  0.02  0.00  0.02  0.02 -0.01  0.01  0.00  0.01  0.01  0.00
## O2          0.01 -0.02 -0.03 -0.01  0.00  0.02  0.01  0.00  0.00  0.02 -0.01
## O3          0.00  0.02  0.01  0.03  0.02  0.00  0.01  0.01 -0.01 -0.01  0.00
## O4          0.01  0.03  0.01  0.02  0.01  0.03  0.03  0.02 -0.02  0.00 -0.01
```

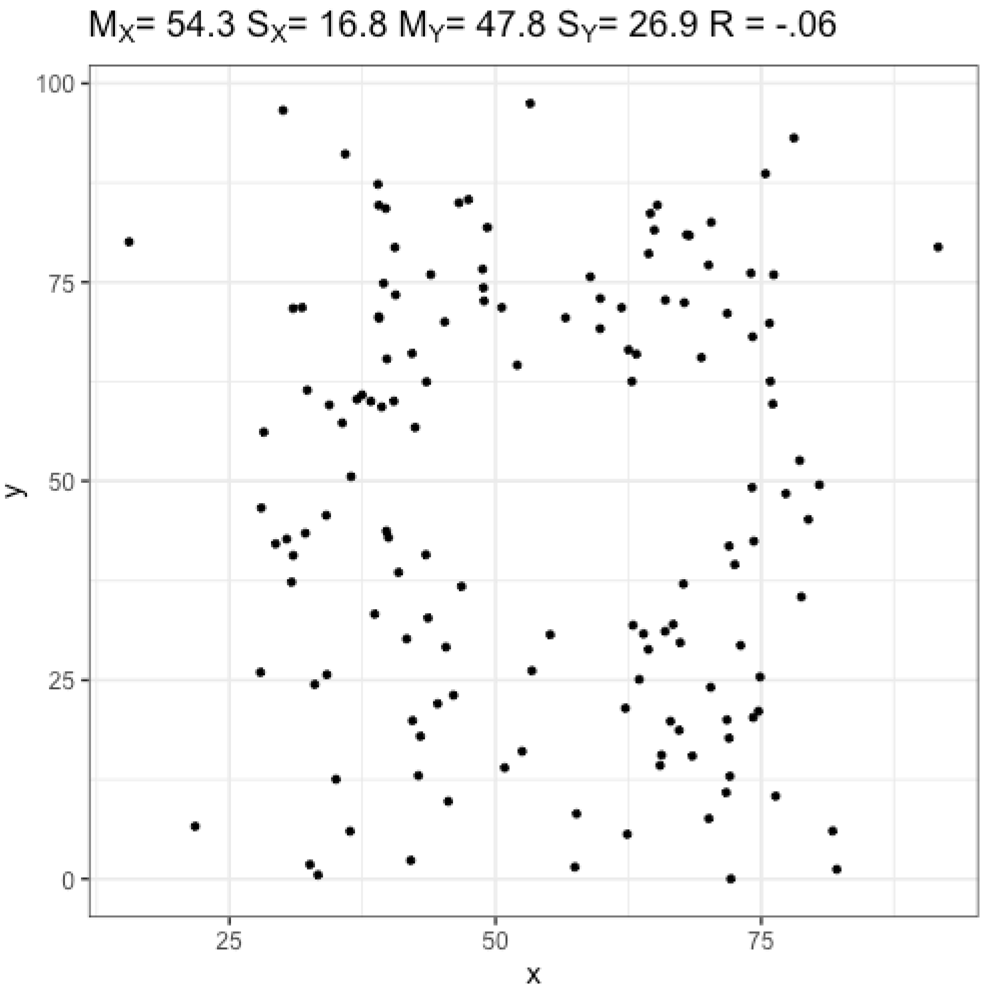
Visualizing correlations

For a single correlation, best practice is to visualize the relationship using a scatterplot. A best fit line is advised, as it can help clarify the strength and direction of the relationship.

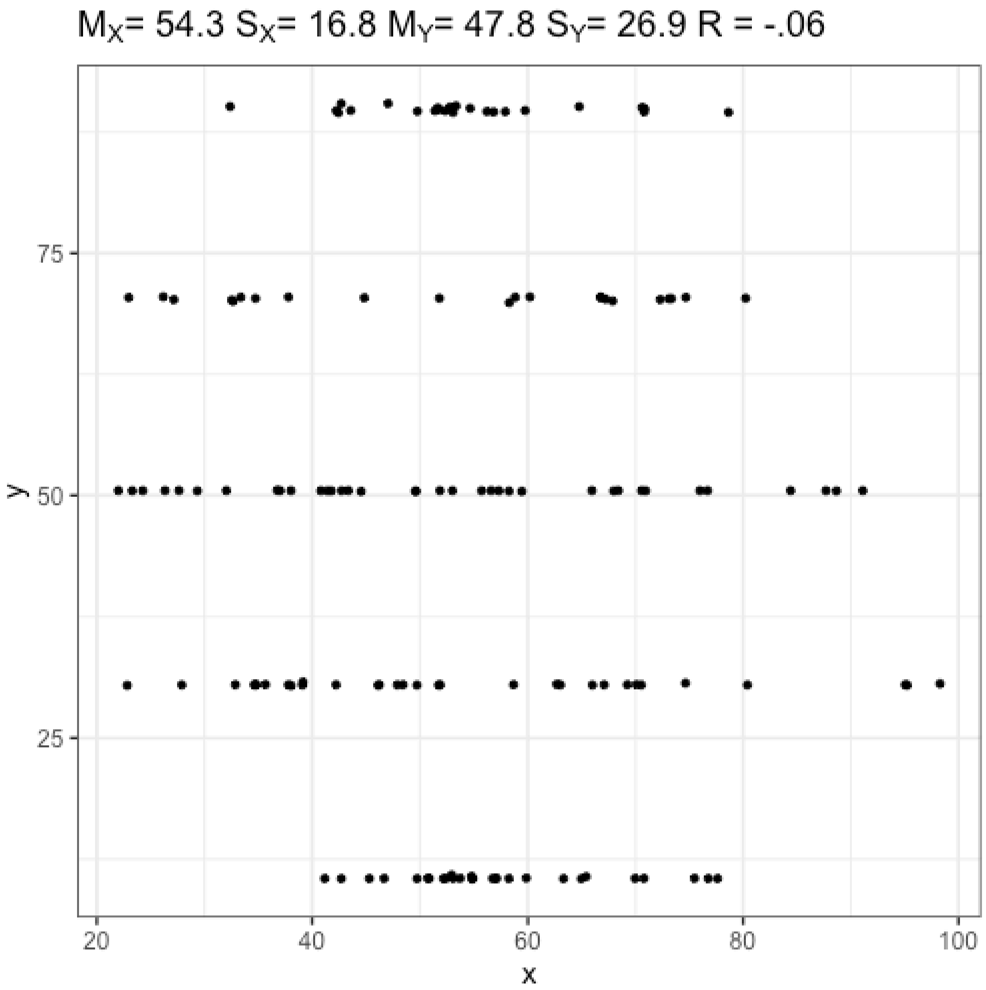
http://guessthecorrelation.com/

Interpreting Correlations

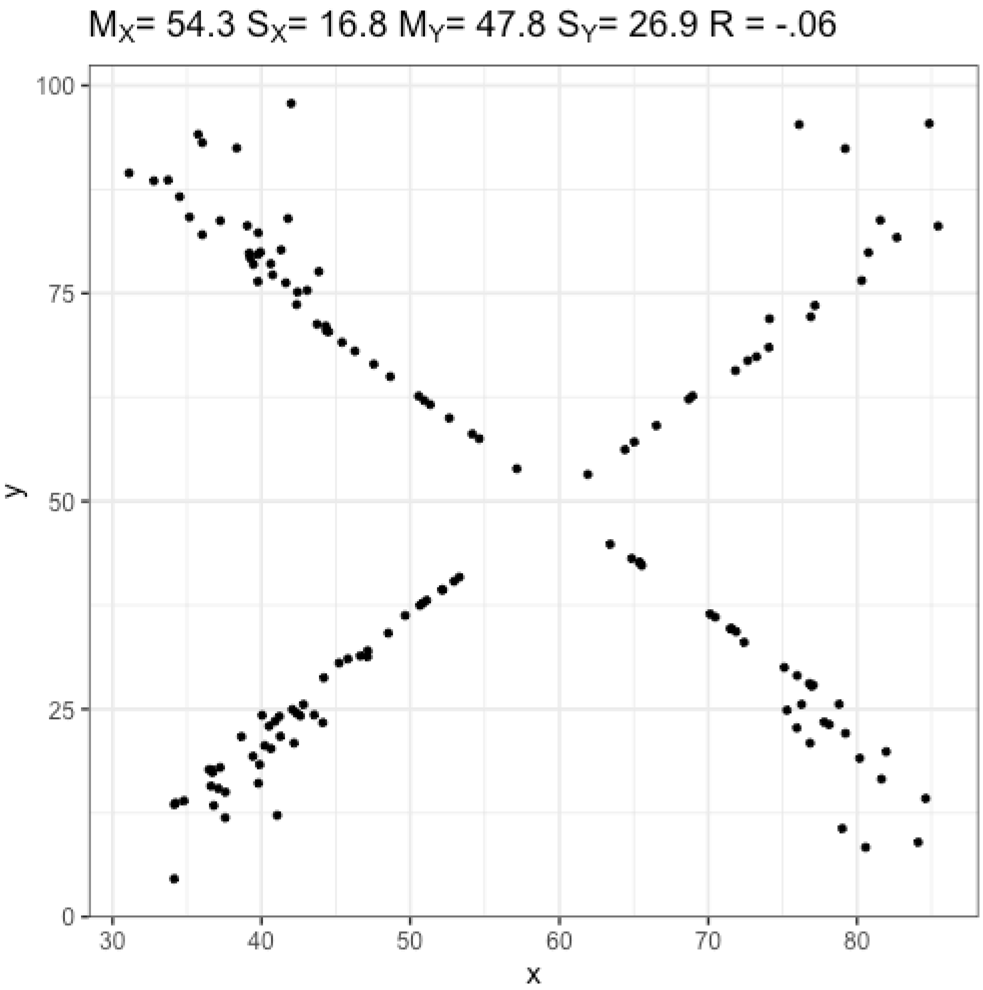
Correlations Can Lie



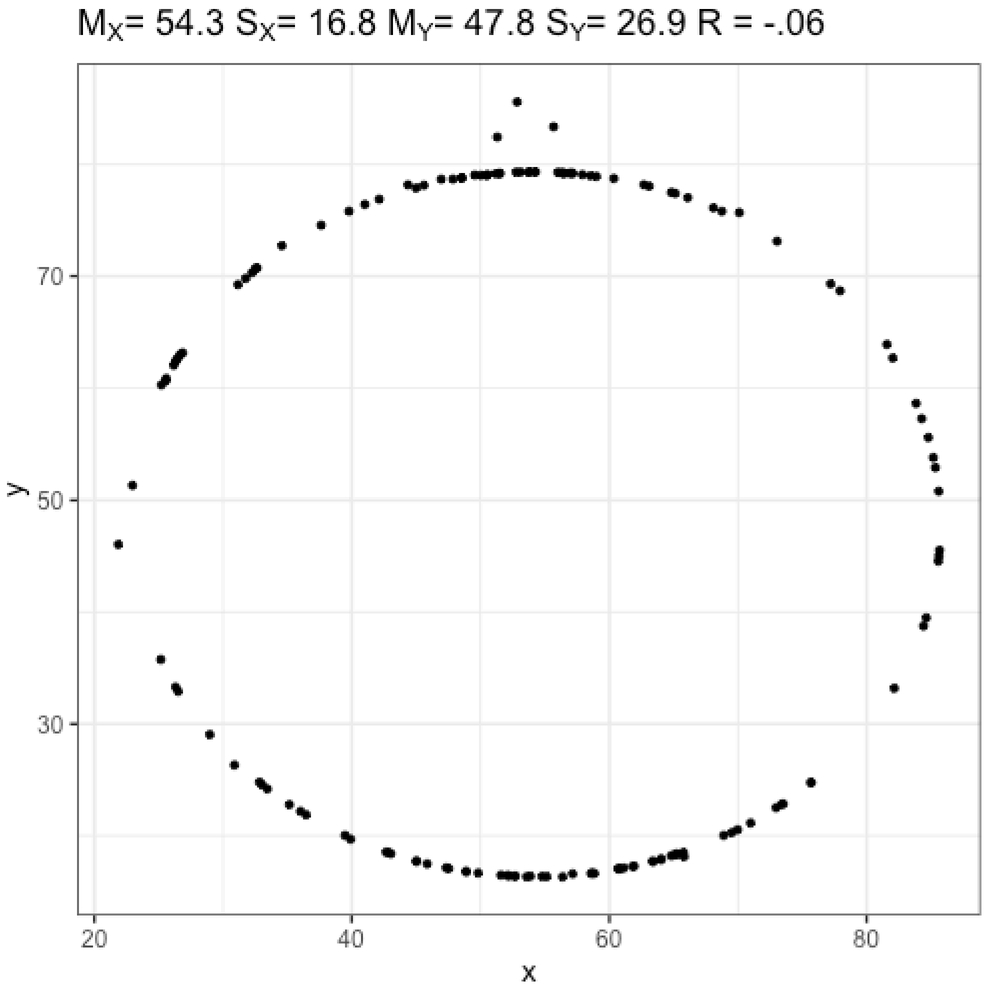
Correlations Can Lie



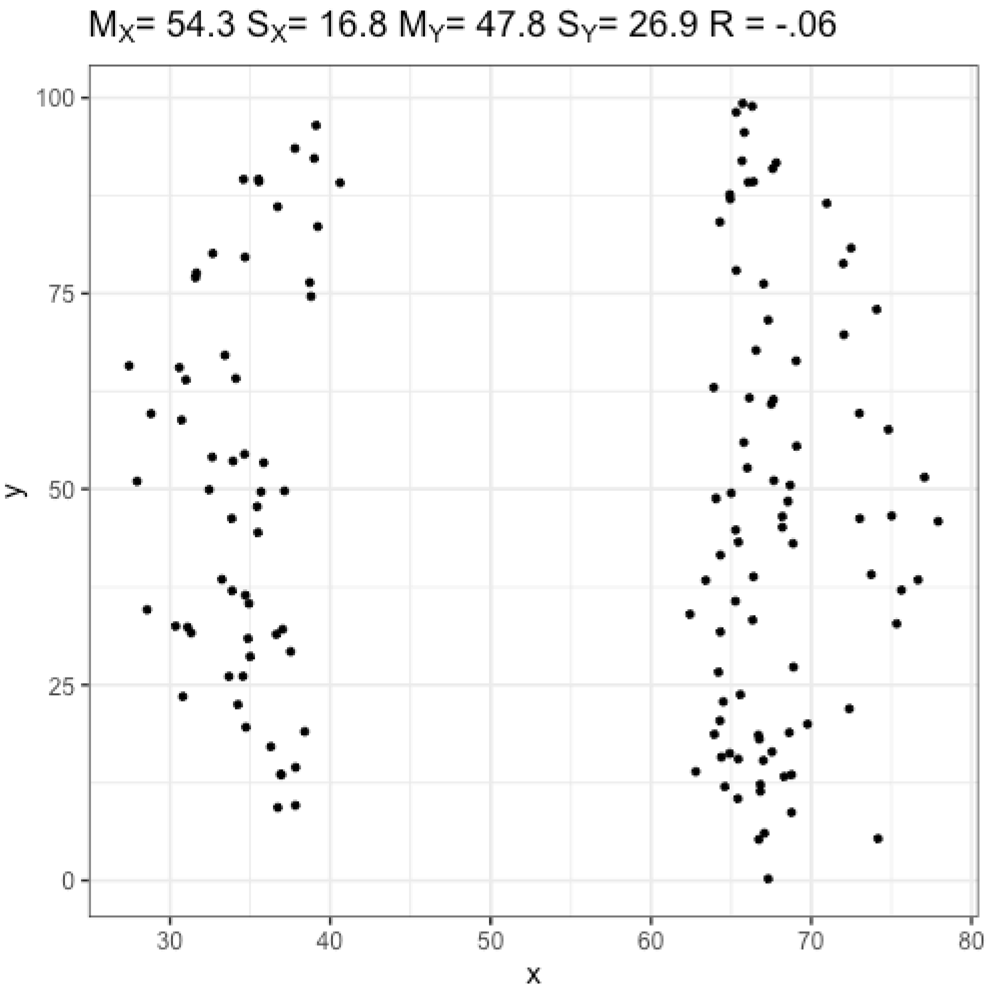
Correlations Can Lie



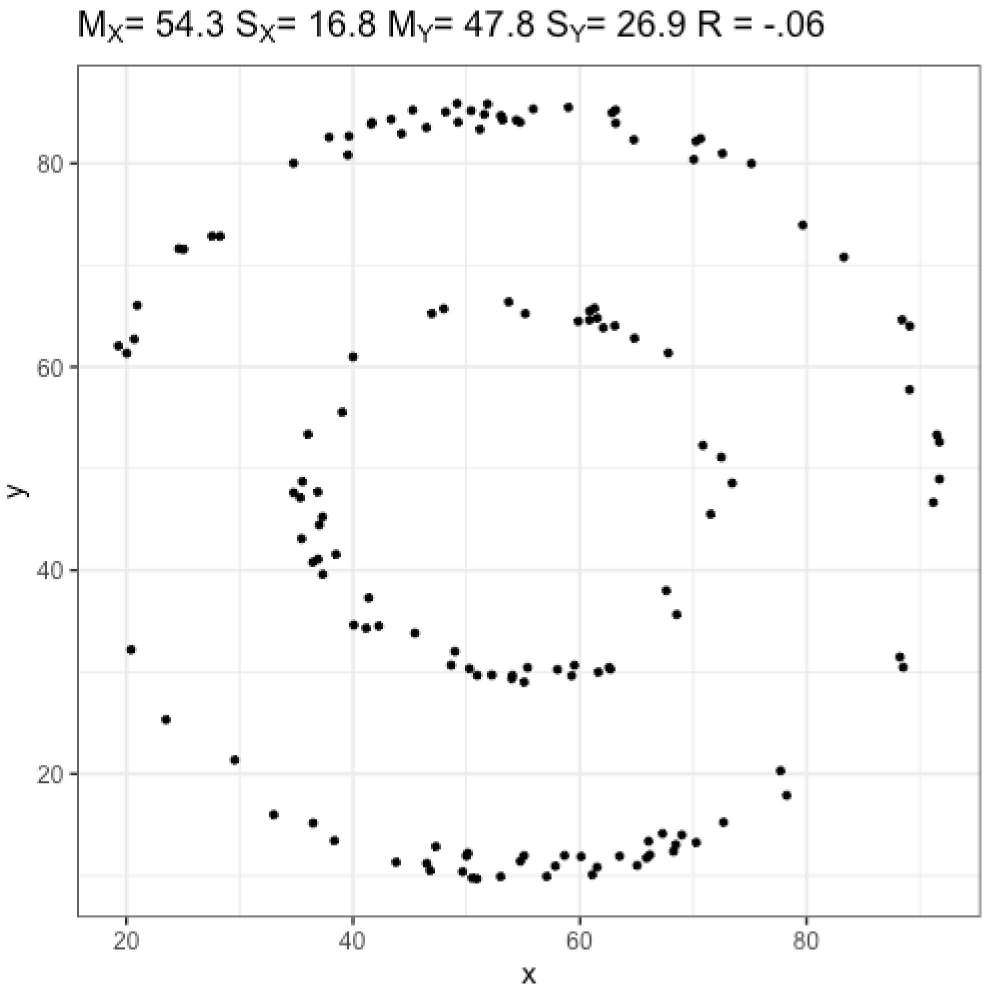
Correlations Can Lie



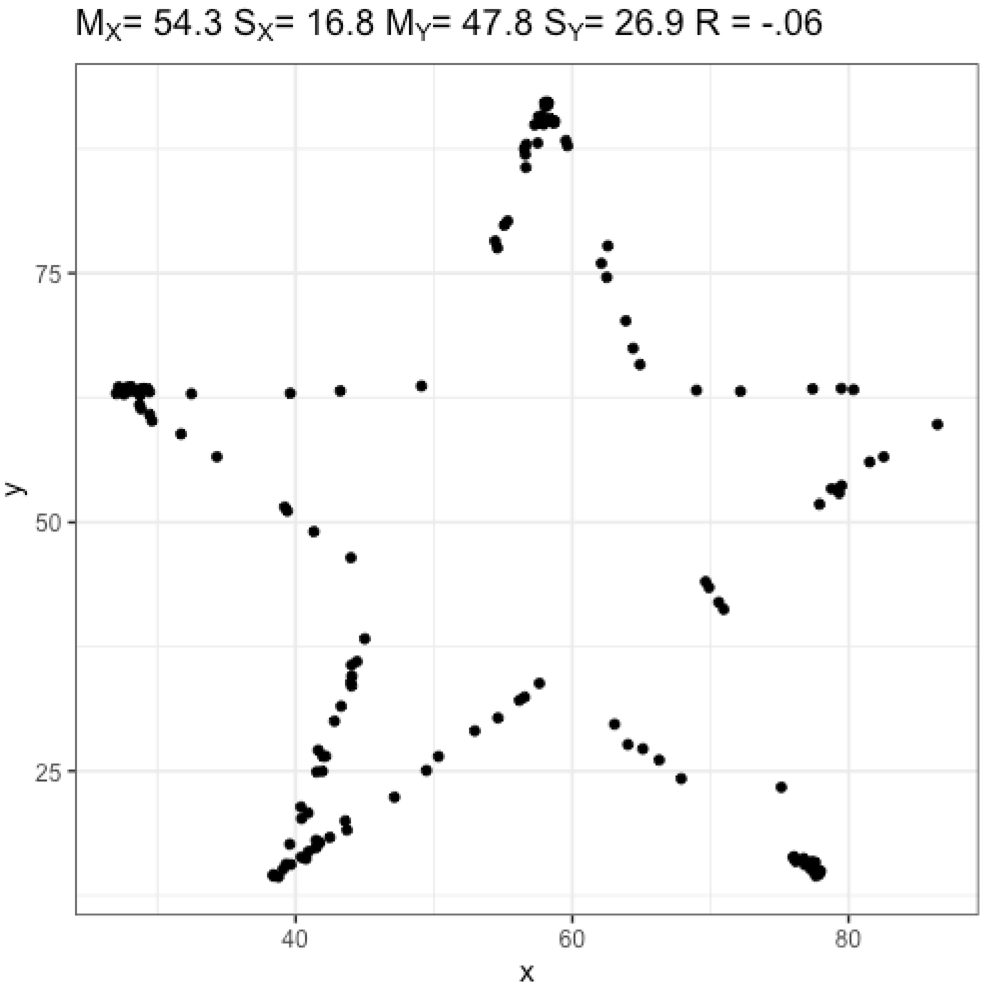
Correlations Can Lie



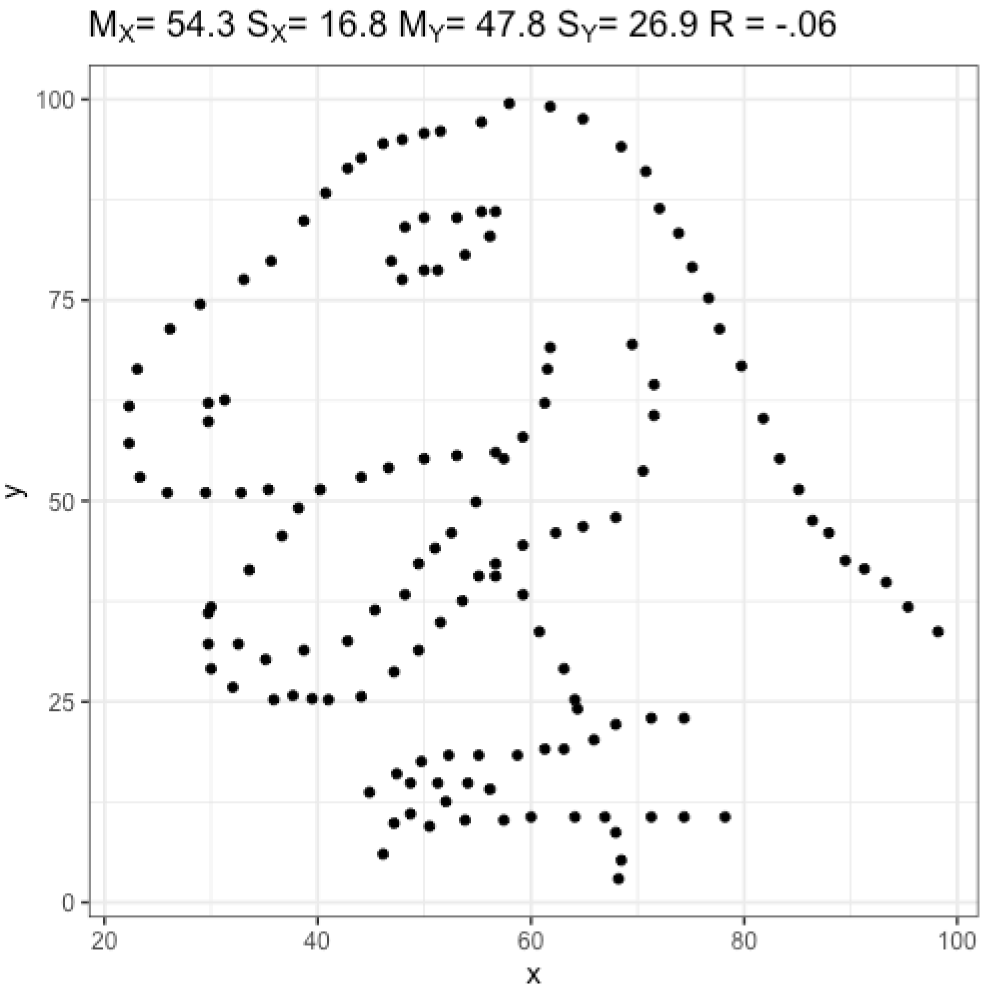
Correlations Can Lie



Correlations Can Lie



Correlations Can Lie



M_X= 54.3 S_X= 16.8 M_Y= 47.8 S_Y= 26.9 R = -.06

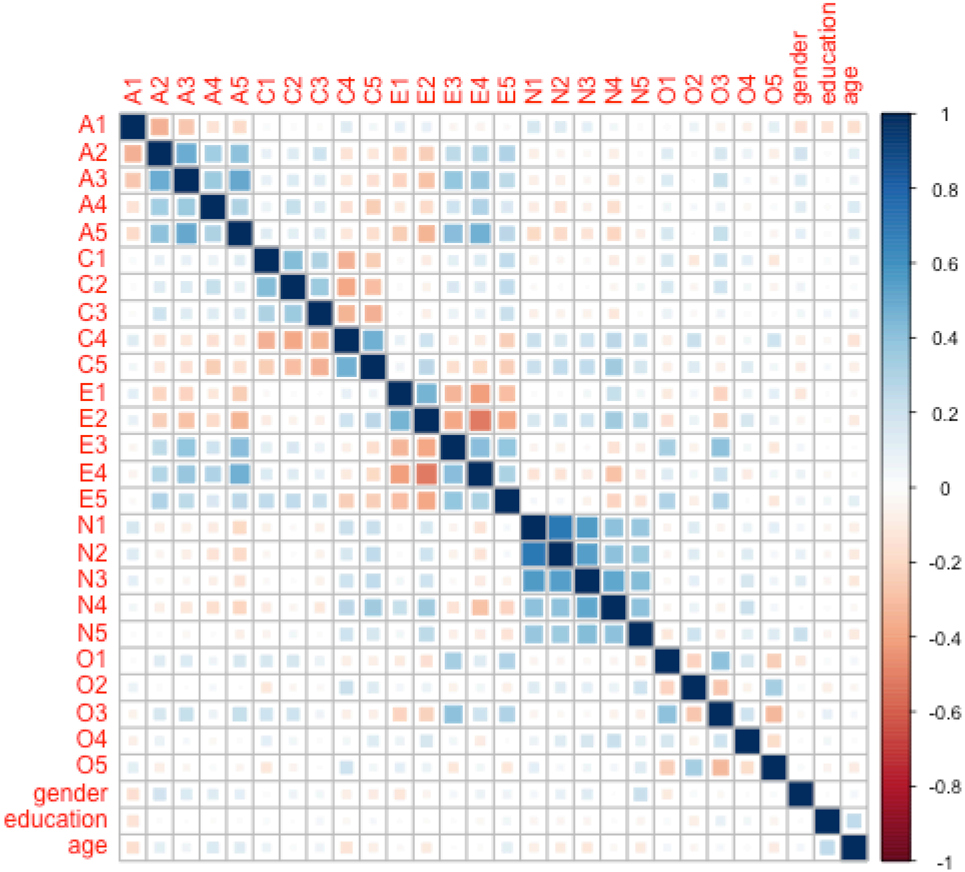
Visualizing correlation matrices

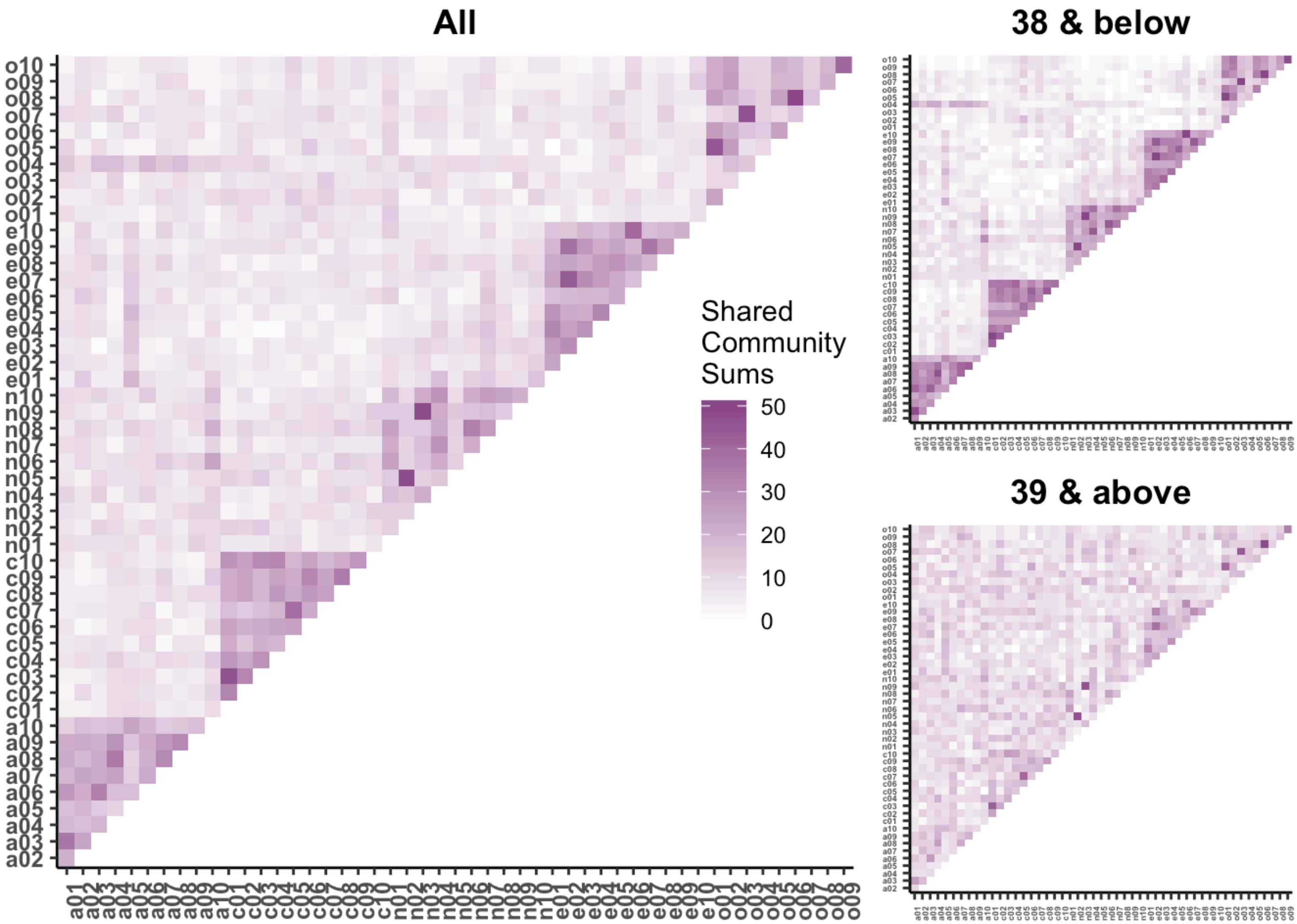
A single correlation can be informative; a correlation matrix is more than the sum of its parts.

Correlation matrices can be used to infer larger patterns of relationships. You may be one of the gifted who can look at a matrix of numbers and see those patterns immediately. Or you can use **heat maps** to visualize correlation matrices.

```
library(corrplot)
```

```
corrplot(cor(bfi, use = "pairwise"), method = "square")
```





Beck, Condon, & Jackson, 2019

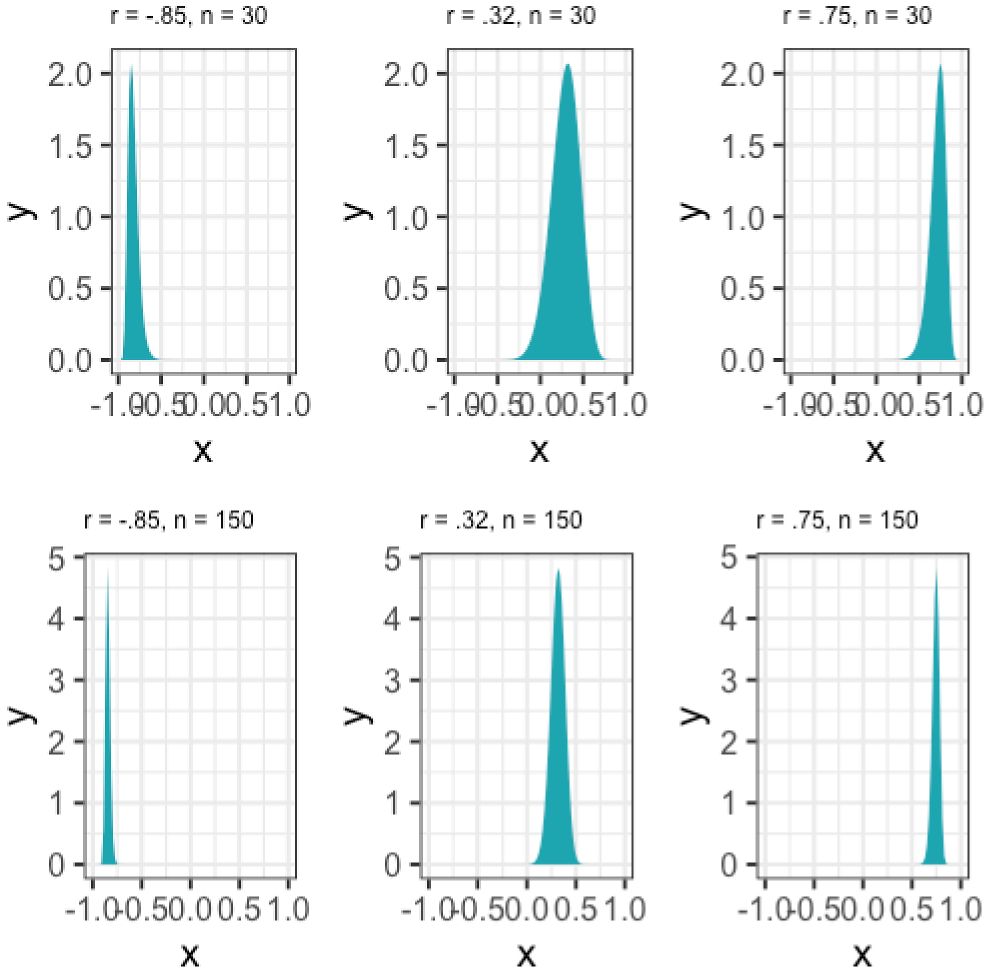
What can we do with correlations?

- Descriptive statistic; describing the strength of association/relationship
- Inferential statistic + hypothesis testing:
    - Is a correlation significantly different from 0?
    - Is a correlation significantly different from a different number? **
    - Construct a confidence interval around our correlation **
    - Are two correlations significantly different from one another? **
    - Used as effect size measure in power calculations

** Need to use a Fisher r to z' transformation

Fisher's r to z' transformation

If we want to make calculations based on $\rho \neq 0$ then we will run into a skewed sampling distribution.

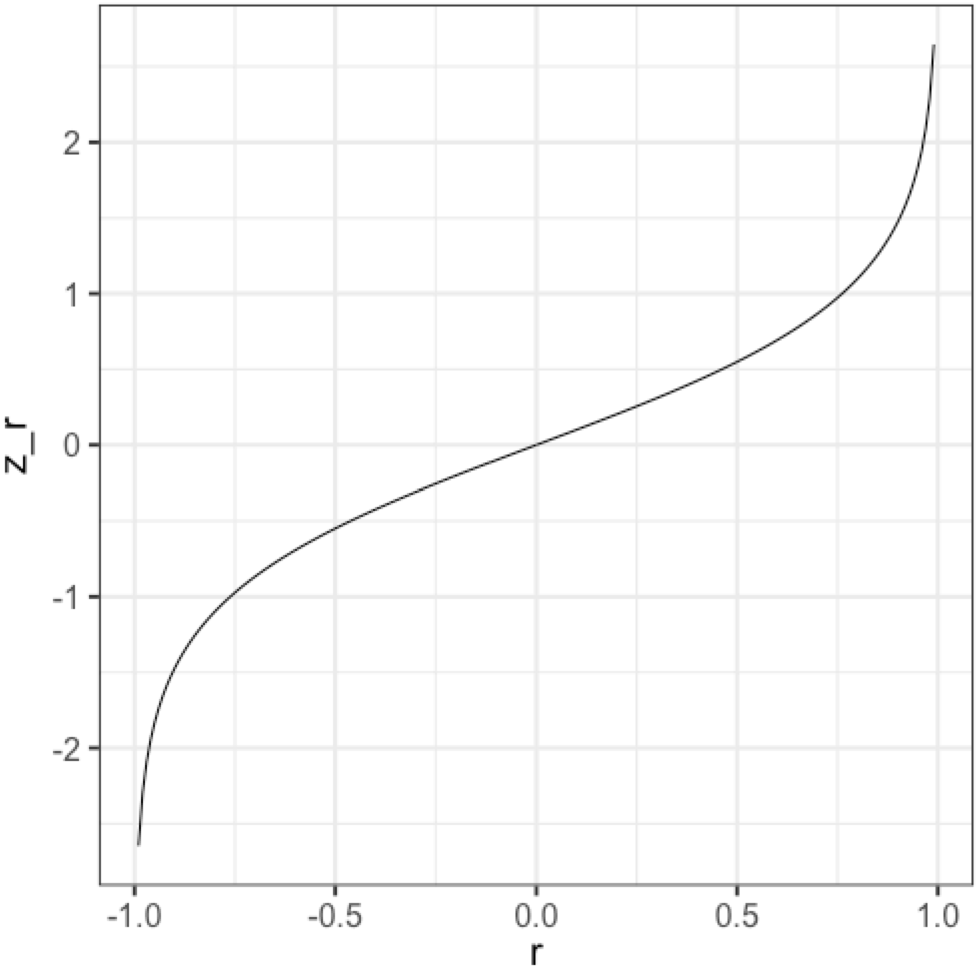


Fisher's r to z' transformation

- Skewed sampling distribution will rear its head when:
  - $H_0 : \rho \neq 0$
  - Calculating confidence intervals
  - Testing two correlations against one another
- r to z':

$$z^{'} = \frac{1}{2} ln \frac{1 + r}{1 - r}$$

Fisher’s r to z’ transformation



How to do in R

```r
library(psych)
fisherz(r)
fisherz2r(z)
```

Use when…

- Is correlation different from a number other than 0?
- Are 2 correlations different from one another?
- Making a confidence interval around a correlation

Statistical Significance

$\neq$

Practical Significance

What is the size of the correlation?

- Chemotherapy and breast cancer survival? (.03)
- Batting ability and hit success on a single at bat? (.06)
- Antihistamine use and reduced sneezing/runny nose? (.11)
- Combat exposure and PTSD? (.11)
- Ibuprofen on pain reduction? (.14)
- Gender and weight? (.26)
- Therapy and well being? (.32)
- Observer ratings of attractiveness? (.39)
- Gender and arm strength? (.55)
- Nearness to equator and daily temperature for U.S.? (.60)

Special cases of the Pearson correlation

- **Spearman correlation coefficient**
  - Applies when both X and Y are ranks (ordinal data) instead of continuous
  - Denoted ρ by your textbook, although I prefer to save Greek letters for population parameters.
- **Point-biserial correlation coefficient**
  - Applies when Y is binary.
    - NOTE: This is not an appropriate statistic when you artificially dichotomize data.
- **Phi (ϕ) coefficient**
  - Both X and Y are dichotomous.

Next time...

Regression!