201ab Quantitative methods non-linear Transformations

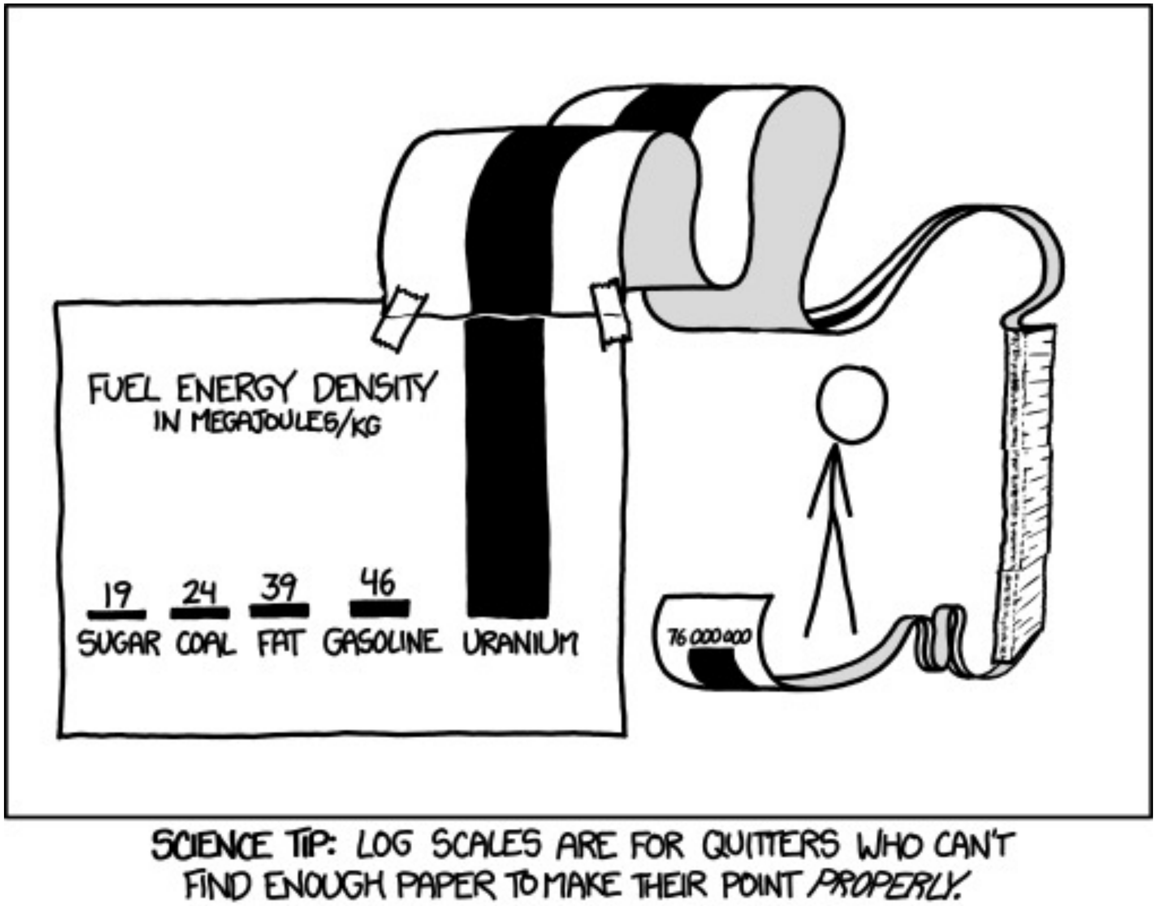
Linearly transforming variables: w' = a*w + b

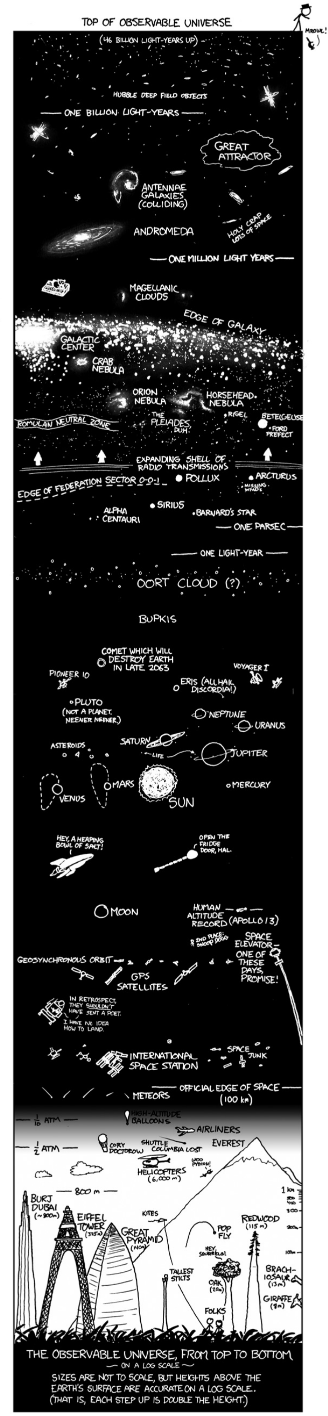
- Centering: X' = X-mean(X)
  makes the intercept mean: Y value at average X
- Z scoring: X' = (X-mean(X))/sd(X)
  also makes the slope mean: change in Y/sd change in X
- Pick real units of X that are of the same order of magnitude as the sd of X.
- Scale dependent variable (Y' = Y*k)
  to make the numerical values of slope and intercept be of a more manageable magnitude

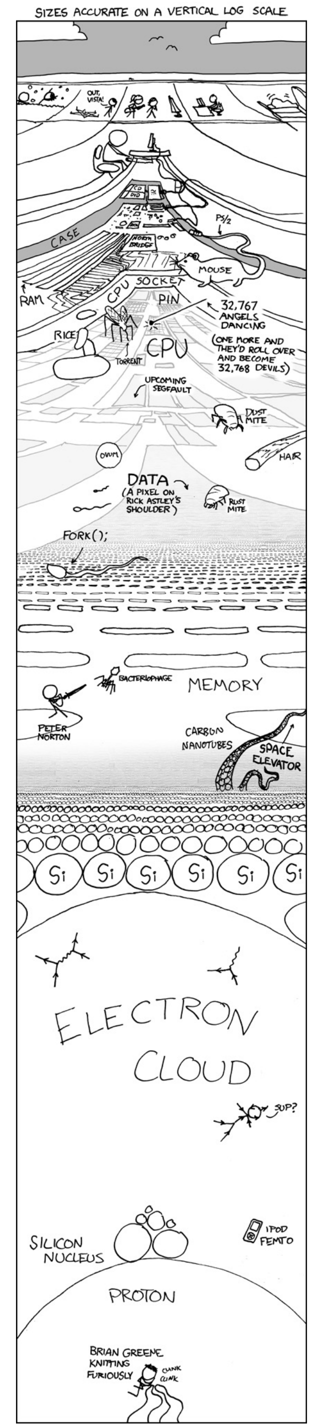
There will be some tradeoffs, and there isn't one 'right' answer (depends on question!) but a bit of scale/unit optimization will help a lot.

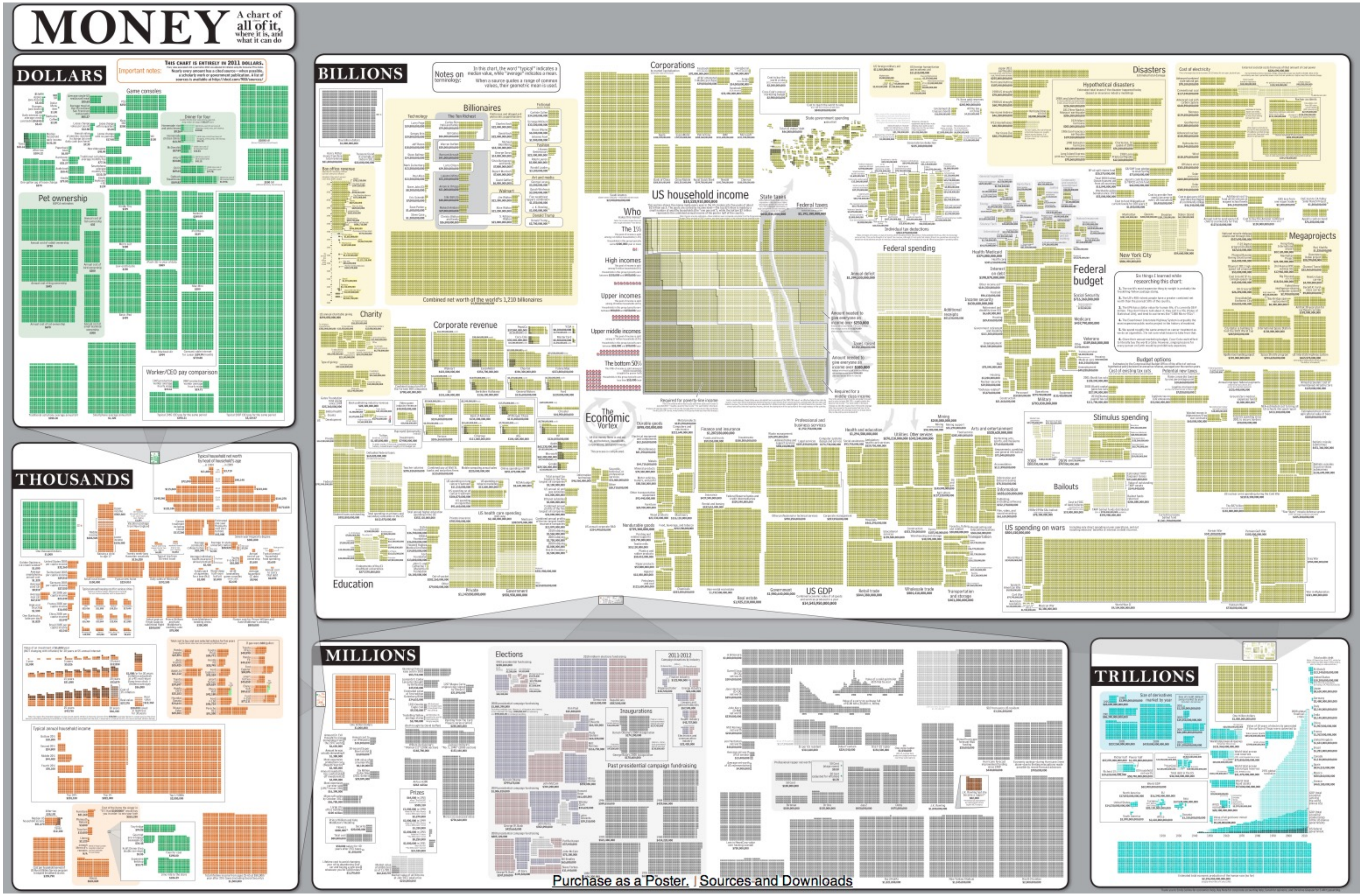
Net worth

• Musk	$281B
• Bezos	$201B
• Gates	$138B
• Buffett	$104B
• Zuckerberg	$121B
• {Alice,Jim,Rob} Walton	$68B
• Marian Ilitch	$4.4B
• Oprah Winfrey	$2.6B
• Lebron James	$850M
• T-Swift	$370M
• Bottom 99%	below $12m
• Median	$125k


FUEL ENERGY DENSITY
IN MEGAJOULES/KG
19
24
39
46
SUGAR
COAL
FAT
GASOLINE
URANIUM
76 000 000
SCIENCE TIP: LOG SCALES ARE FOR QUITTERS WHO CAN'T FIND ENOUGH PAPER TO MAKE THEIR POINT PROPERLY.


TOP OF OBSERVABLE UNIVERSE
ONE BILLION LIGHT-YEARS
GREAT ATTRACTOR
ANTENNAE GALAXIES (COLLIDING)
ANDROMEDA
ONE MILLION LIGHT YEARS
MAGELLANIC CLOUDS
EDGE OF GALAXY
GALACTIC CENTER
CRAB NEBULA
ORION NEBULA
HORSEHEAD NEBULA
RIGEL
BETELGEUSE
ROMULAN NEUTRAL ZONE
EXPANDING SHELL OF RADIO TRANSMISSIONS
EDGE OF FEDERATION SECTOR 0-0-1
POLLUX
ARCTURUS
SIRIUS
ALPHA CENTAURI
BARNARD'S STAR
ONE PARSEC
ONE LIGHT-YEAR
OORT CLOUD (?)
BUPKIS
COMET WHICH WILL DESTROY EARTH IN LATE 2063
PIONEER 10
ERIS
PLUTO
NEPTUNE
URANUS
SATURN
JUPITER
ASTEROIDS
MARS
VENUS
SUN
MERCURY
MOON
GEOSYNCHRONOUS ORBIT
GPS SATELLITES
INTERNATIONAL SPACE STATION
OFFICIAL EDGE OF SPACE (100 km)
METEORS
AIRLINERS
EVEREST
HELICOPTERS
BURJ DUBAI
EIFFEL TOWER
GREAT PYRAMID
REDWOOD
THE OBSERVABLE UNIVERSE, FROM TOP TO BOTTOM
ON A LOG SCALE
SIZES ARE NOT TO SCALE, BUT HEIGHTS ABOVE THE EARTH'S SURFACE ARE ACCURATE ON A LOG SCALE.
(THAT IS, EACH STEP UP IS DOUBLE THE HEIGHT.)


SIZES ACCURATE ON A VERTICAL LOG SCALE
CASE
PS/2
MOUSE
RAM
CPU
SOCKET
PIN
32,767 ANGELS DANCING
(ONE MORE AND THEY'D ROLL OVER AND BECOME 32,768 DEVILS)
RICE
UPCOMING SEGFAULT
DUST MITE
HAIR
DATA (A PIXEL ON RICK ASTLEY'S SHOULDER)
FORK();
MEMORY
PETER NORTON
CARBON NANOTUBES
SPACE ELEVATOR
Si
ELECTRON CLOUD
SILICON NUCLEUS
PROTON
BRIAN GREENE KNITTING FURIOUSLY


MONEY
A chart of all of it, where it is, and what it can do
DOLLARS
THOUSANDS
BILLIONS
MILLIONS
TRILLIONS
Billionaires
Corporations
US household income
Federal spending
Federal budget
Disasters
Megaprojects
Corporate revenue
Charity
Education
US GDP
US spending on wars
Stimulus spending
Bailouts
Elections
Prizes
Pet ownership
Game consoles
Purchase as a Poster. | Sources and Downloads

The log transform

- Why use the log transform?
- For visualization: Some measures vary over orders of magnitude and are simply unmanageable on a linear scale.

- For your statistical model: usually those measures that vary over many orders of magnitude do not have *linear* relationships with other measures, and the kinds of non-linear relationships they have are well captured with a logarithm.

Transformations

- Log transform
  - Logarithms
  - Log transforming response variables
  - Log transforming predicting variables
  - Log transforming response and predicting variables
- Logit transform (covered in logistic regression)
  - Logit and logistic transformations (inverses of each other)
  - Logit(y) ~ x
  - Y~logit(x) ?

Exponents and Logarithms

a to the power of b

$$\overbrace{a^b}^{\text{a to the power of b}} = \underbrace{a * a * a * \ldots * a}_{b\text{–}times}$$

What do you get if you multiply a times itself b times.

$$\underbrace{\log_a}_{\text{Log “base a”}}[a^b] = b$$

How many times do you need to multiply a times itself to get this number

Log “base a”

If you don't like standard notation: https://www.youtube.com/watch?v=sULa9Lc4pck

The log transform

$$5^6 = \underbrace{5*5*5*5*5*5}_{6-times} = 15625$$

```
5^6
15625
```

```
log(15625,5)
6
```

$$\log_5[15625] = 6$$

- Common bases for logs
  - Log2 (useful for binary things, e.g., bits in memory)

```
2^c(    1,    2,   3,   4,    5,    6,   7,    8,    9,   10 )
        2     4    8   16    32    64   128   256   512  1024
```

  - Log base e ('natural log') e = 2.718282.
    (arises from continuous compounding)

```
exp(1)^c(   1,  2,   3,    4,     5,     6,      7,      8,      9,      10    )
           2.7 7.4 20.1 54.6 148.4 403.4 1096.6 2981.0 8103.1 22026.5
```

  - Log base 10 (very intuitive – my preferred base!)

```
10^c(  1,  2,    3,     4,      5,       6,        7,         8,          9,          10          )
      10 100 1000 10000 100000 1000000 10000000 100000000 1000000000 10000000000
```

The log transform

- Exponents

$$a^n a^m = a^{n+m}$$

$$\frac{a^n}{a^m} = a^{n-m} = \frac{1}{a^{m-n}}$$

$$\left(a^n\right)^m = a^{nm}$$

$$a^0 = 1, \quad a \neq 0$$

$$(ab)^n = a^n b^n$$

$$\left(\frac{a}{b}\right)^n = \frac{a^n}{b^n}$$

$$a^{-n} = \frac{1}{a^n}$$

$$\frac{1}{a^{-n}} = a^n$$

$$\left(\frac{a}{b}\right)^{-n} = \left(\frac{b}{a}\right)^n = \frac{b^n}{a^n}$$

$$a^{\frac{n}{m}} = \left(a^{\frac{1}{m}}\right)^n = \left(a^n\right)^{\frac{1}{m}}$$

- Logarithms

Definition

$y = \log_b x$ is equivalent to $x = b^y$

Example

$\log_5 125 = 3$ because $5^3 = 125$

Logarithm Properties

$$\log_b b = 1 \qquad \log_b 1 = 0$$

$$\log_b b^x = x \qquad b^{\log_b x} = x$$

$$\log_b\left(x^r\right) = r \log_b x$$

$$\log_b\left(xy\right) = \log_b x + \log_b y$$

$$\log_b\left(\frac{x}{y}\right) = \log_b x - \log_b y$$

$\log_a(x) = \log_b(x) / \log_b(a)$

Log-math Practice

1) $\log_{10}(x) = 4*\log_{10}(y) + 2$. What is y=?
2) $\log_{10}(x) = 4*y + 2$. What is $\log_2(x)$=?
3) $\log_{10}(y) = 0.3*x + 3$
   how does y change when x increases by +2? by *2?
4) $\log_{10}(y) = 0.3*\log_{10}(x) + 3$
   how does y change when x increases by +2? by *2?
5) $y = 0.3*\log_{10}(x) + 3$
   how does y change when x increases by +2? by *2?

Reasoning about regressions with log transforms requires thinking about exponents and logarithms. If you are rusty on exponents and logarithms, please refresh.
Khan academy: https://www.khanacademy.org/math/algebra-home/alg-exp-and-log
Paul's Algebra notes: https://tutorial.math.lamar.edu/Classes/Alg/Alg.aspx
Paul's Online Notes cheatsheet: https://tutorial.math.lamar.edu/getfile.aspx?file=B,30,N

Transformations

- Linear transformations
  - Predicting variables
  - Response variables
- Log transform
  - Logarithms
  - Log transforming response variables
  - Log transforming predicting variables
  - Log transforming response and predicting variables
- Logit transform (maybe today, maybe later in logistic)
  - Logit and logistic transformations (inverses of each other)
  - Logit(y) ~ x
  - Y~logit(x) ?

The log transform

- Why use the log transform?

- Some measures vary over orders of magnitude and are simply unmanageable on a linear scale

- Some measures are not sums of their predictors, but products. (often yielding measures varying over orders of magnitude)
  - A log transform makes them additive
    $\log(x*y) = \log(x) + \log(y)$

log-transforming response variable

- Instead of:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \varepsilon_i$$

- We do:

$$\log_{10}(Y_i) = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \varepsilon_i$$

- Therefore:

$$Y_i = 10^{\left[\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \varepsilon_i\right]}$$

$$Y_i = 10^{\beta_0} 10^{\beta_1 X_{1i}} 10^{\beta_2 X_{2i}} 10^{\varepsilon_i}$$

- So what does a slope of $B_1 = 2$ mean?

log-transforming response variable

$$\log_{10}(Y_i) = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \varepsilon_i$$

- Therefore:

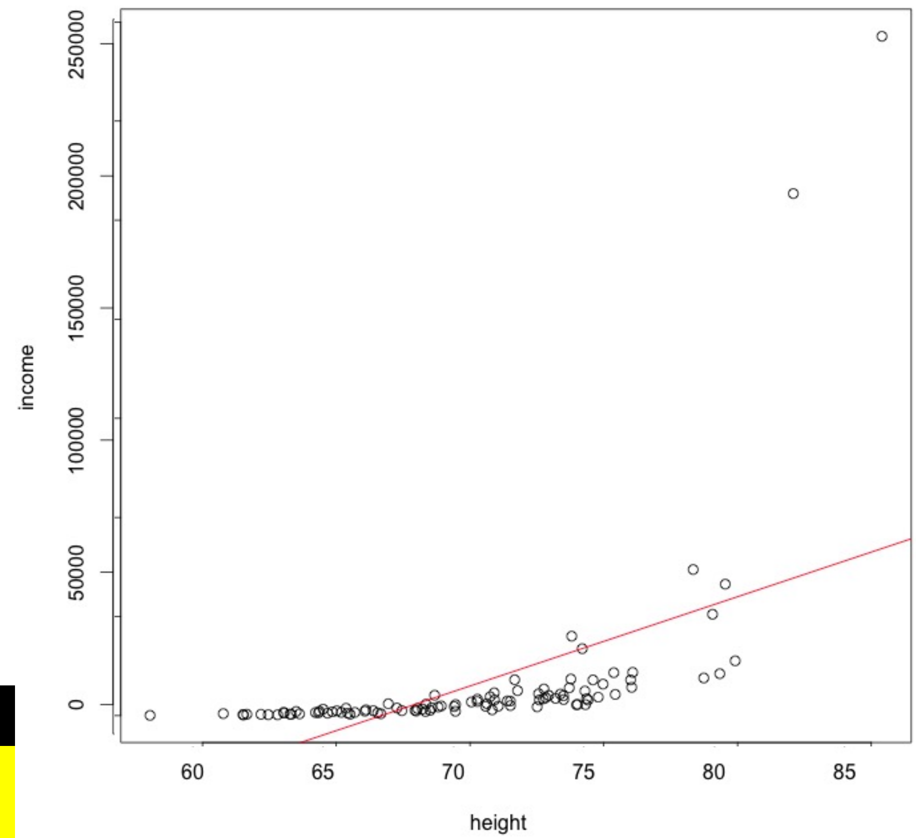
$$Y_i = 10^{\left[\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \varepsilon_i\right]}$$

$$Y_i = 10^{\beta_0} 10^{\beta_1 X_{1i}} 10^{\beta_2 X_{2i}} 10^{\varepsilon_i}$$

- So what does a slope of $B_1 = 2$ mean?
  - For every unit increase of X1 (all else equal) the base-10 log of Y goes up by 2.
  - For every unit increase of X1 (all else equal) Y goes up by a **factor of10^2=100!**

Log regression example

- Income vs height



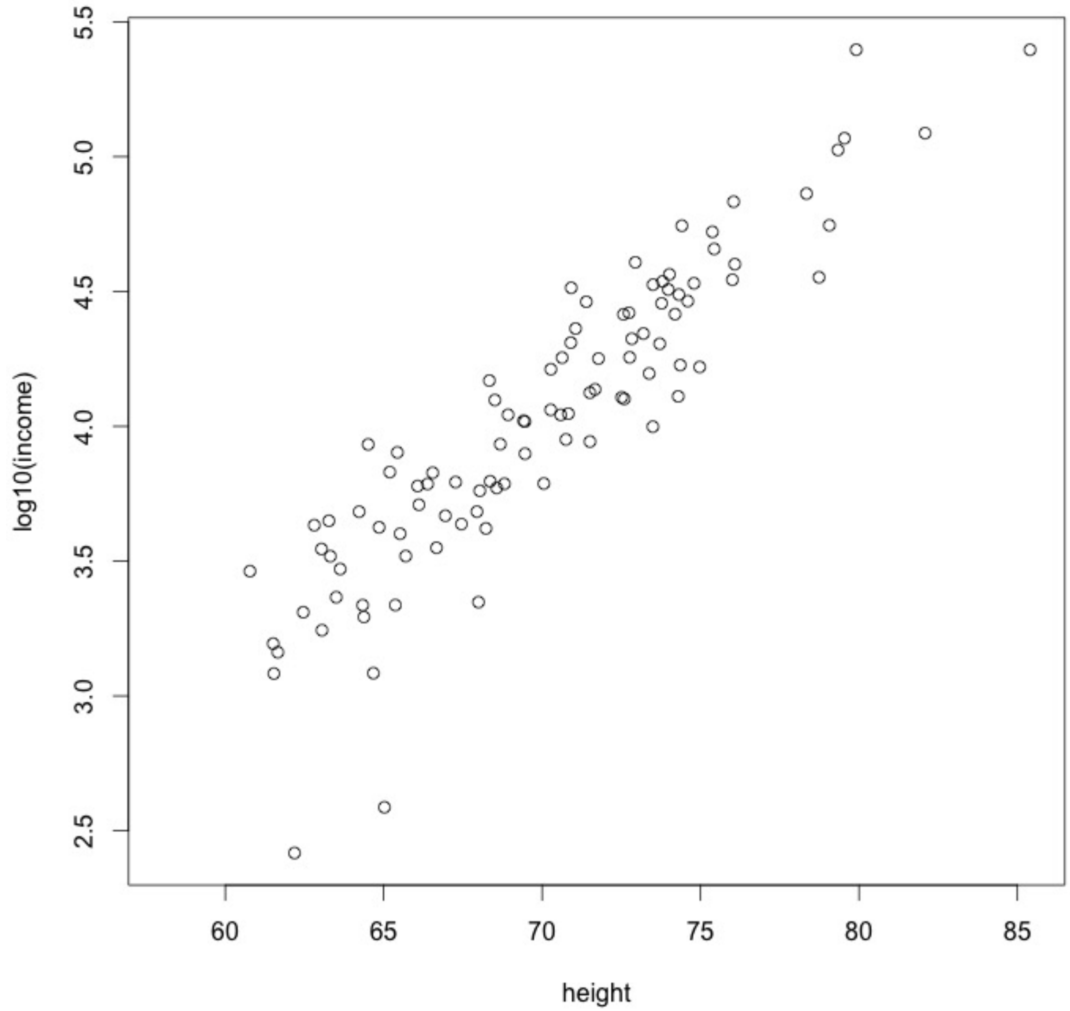

```
summary(lm(income~height))
```

```
Residuals:
   Min     1Q Median     3Q    Max
-34607 -15335  -6904   8686 172609

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -351363.2    37988.1  -9.249 5.16e-15 ***
height         5355.1      541.4   9.891  < 2e-16 ***
---
Residual standard error: 28230 on 98 degrees of freedom
Multiple R-squared: 0.4996,	Adjusted R-squared: 0.4945
F-statistic: 97.84 on 1 and 98 DF,  p-value: < 2.2e-16
```

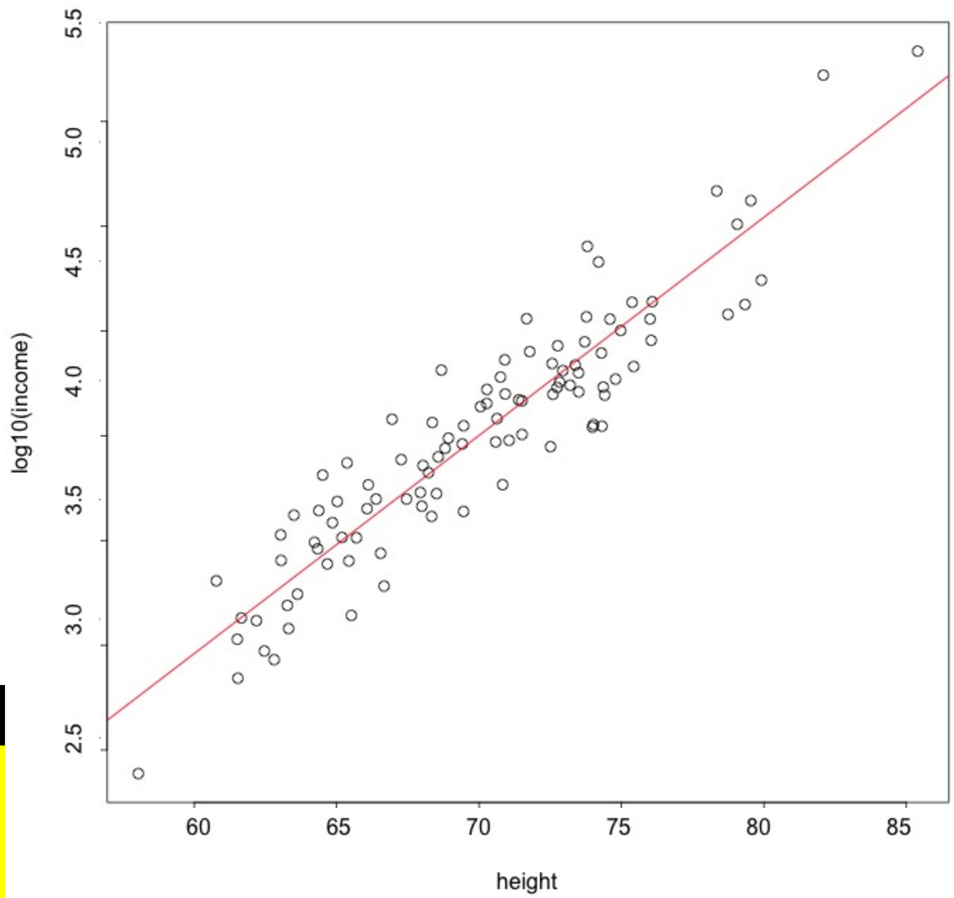
Log regression example

- Income vs height



Log regression example

- Log10(Income) vs height
- What does…
  0.104162 mean?
  -3.29 mean?

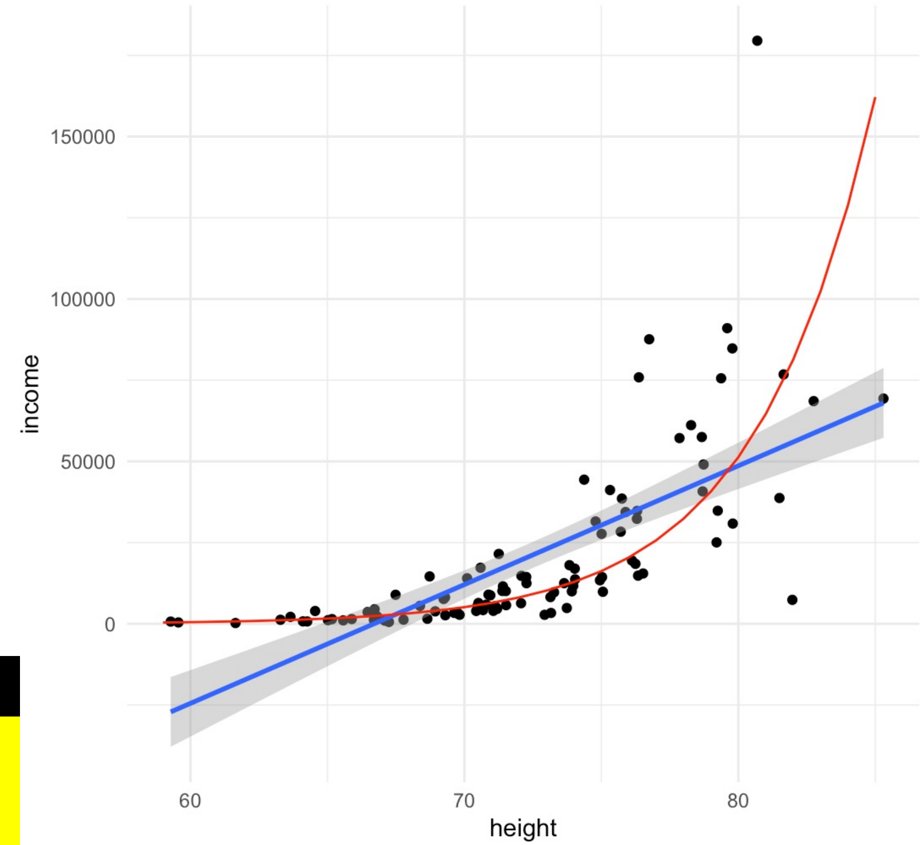



```
summary(lm(log10(income)~height))
Residuals:
      Min        1Q    Median        3Q       Max
-0.404473 -0.137240  0.007002  0.129492  0.507423

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -3.290729   0.294412  -11.18   <2e-16 ***
height       0.104162   0.004196   24.82   <2e-16 ***
---
Residual standard error: 0.2188 on 98 degrees of freedom
Multiple R-squared: 0.8628,    Adjusted R-squared: 0.8614
F-statistic: 616.3 on 1 and 98 DF,  p-value: < 2.2e-16
```

Log regression example

- Log10(Income) vs height
- What does…
  0.104162 mean?
  -3.29 mean?




```
summary(lm(log10(income)~height))

Residuals:
      Min        1Q    Median        3Q       Max
-0.404473 -0.137240  0.007002  0.129492  0.507423

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -3.290729   0.294412  -11.18   <2e-16 ***
height       0.104162   0.004196   24.82   <2e-16 ***
---
Residual standard error: 0.2188 on 98 degrees of freedom
Multiple R-squared: 0.8628,   Adjusted R-squared: 0.8614
F-statistic: 616.3 on 1 and 98 DF,  p-value: < 2.2e-16
```

Log regression example

- Log10(Income) vs height
- What does 0.104162 mean?
  - For every inch taller, log10(income) goes up by 0.1
  - For every inch taller, income goes up by a **factor of** 10^0.1 (1.26).
  - For every inch taller, you will make 26% more
- What does -3.29 mean?
  - At height=0:
    log10(income)=-3.29
    income=10^-3.29
    income=$0.0005

```
summary(lm(log10(income)~height))

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -3.290729   0.294412  -11.18   <2e-16 ***
height       0.104162   0.004196   24.82   <2e-16 ***
```

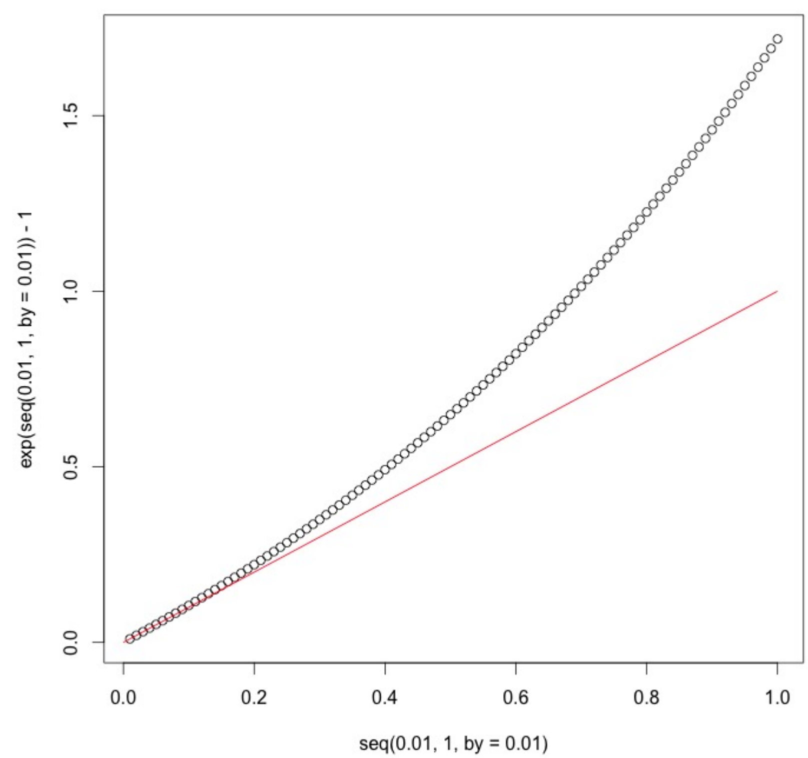
Log transform desiderata

- Which log?
- When to use log transform?
- When not to use it?
- What to do about zeros?
- Confidence intervals with non-linear transforms…

Natural log or log base 10?

- Log base 10 is handy because the predicted y values are easy to interpret.

- Log base e (natural log) is handy because the coefficients are easy to interpret due to small number approximation (a coefficient of 0.05 means a 5% increase per unit x)



When to log transform response variables?

- When effects of predictors and noise are proportional.
  - As arise from various growth processes…
- This often arises when…
  - …response variable is bounded at (and is close to) zero
    Ratios, speed, income, time, height, distance, contrast, sensitivity, etc…
  - …variance scales with mean (Weber noise)
    Estimation of physical properties, spike counts, etc.
- These often co-occur: proportional effects yield proportional errors, variance scaling with mean, bounds at zero…

When not to log transform response variables

- When responses can be negative!
  - Linear!

- When predictors seem to be additive.
  - Linear!

- When you have an upper bound (e.g. proportions) (consider logit, later)

What to do about zeros?

Log(0) is undefined… so if you have zeros, you can't log.

- Option 1: decide that zeros are real, and it would be wrong to coerce them to behave… try something else (maybe Poisson regression)
- Option 2: change zeros to something small (smaller than the smallest non-zero unit), to get them to behave (e.g., population=0? Call that population=1)
- Option 3: change everything by adding a small offset (e.g., pop' = population + 1)

Have a principled reason for choosing small unit, and hope that it doesn't have much of an effect.

Confidence intervals for linearized lm

- Let's say log10(y)~B0+B1*x
  Estimates: B1 = 1, se{B1} = 0.2

- What is the 95% interval on the change in y per unit increase of x?

Confidence intervals for linearized lm

- Let's say log10(y)~B0+B1*x
  Estimates: B1 = 1, se{B1} = 0.2

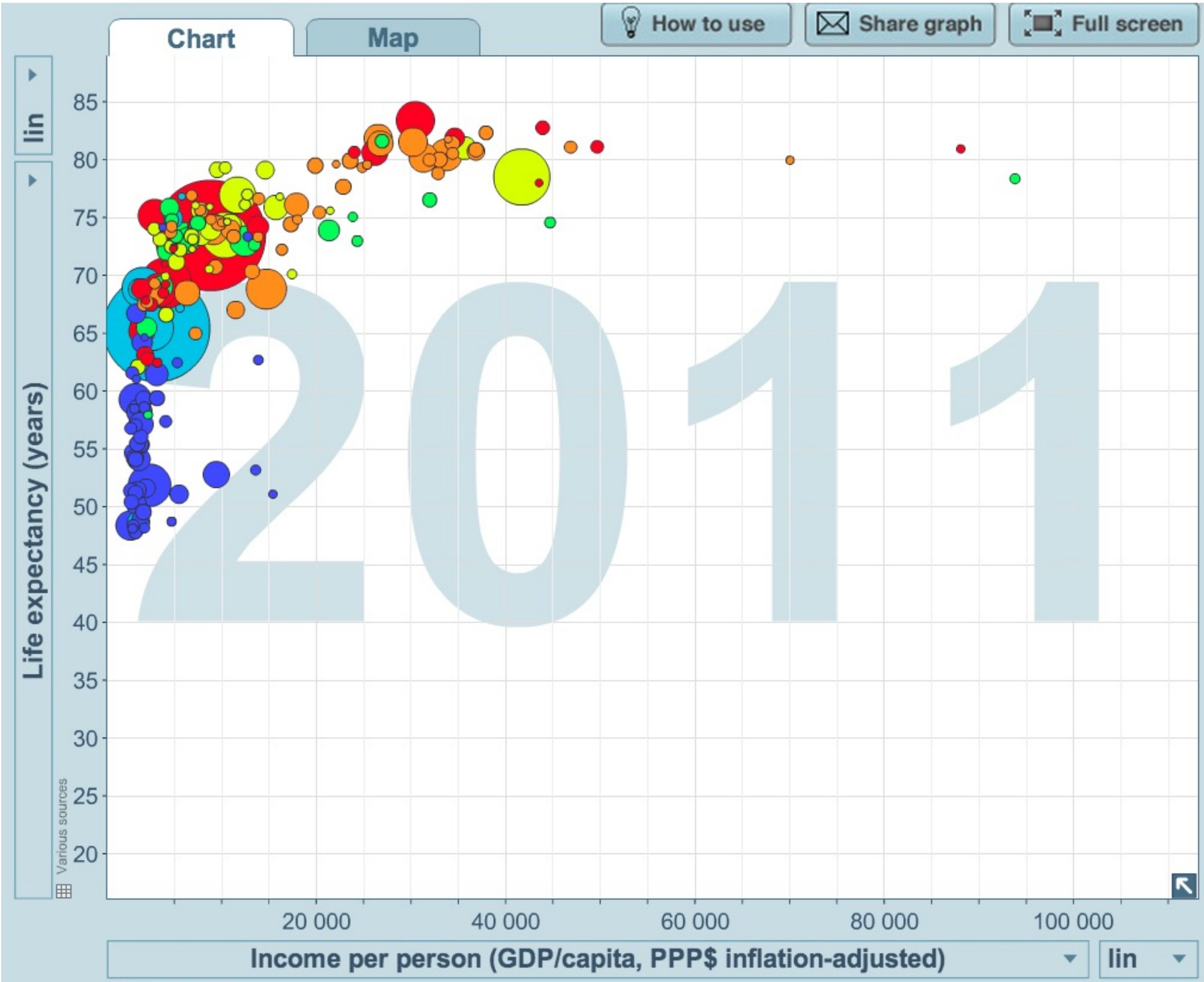
- What is the 95% interval on the change in y per unit increase of x?
  - 95% CI on B1 = 0.6 to 1.4
    (this is change in log10(y) per unit increase of x)
  - 95% CI on proportional change to B1 per unit increase of x:
    10^0.6 to 10^1.4 -> 4 to 25

- Basically: transform *after* obtaining a confidence interval – meaningless to transform before.

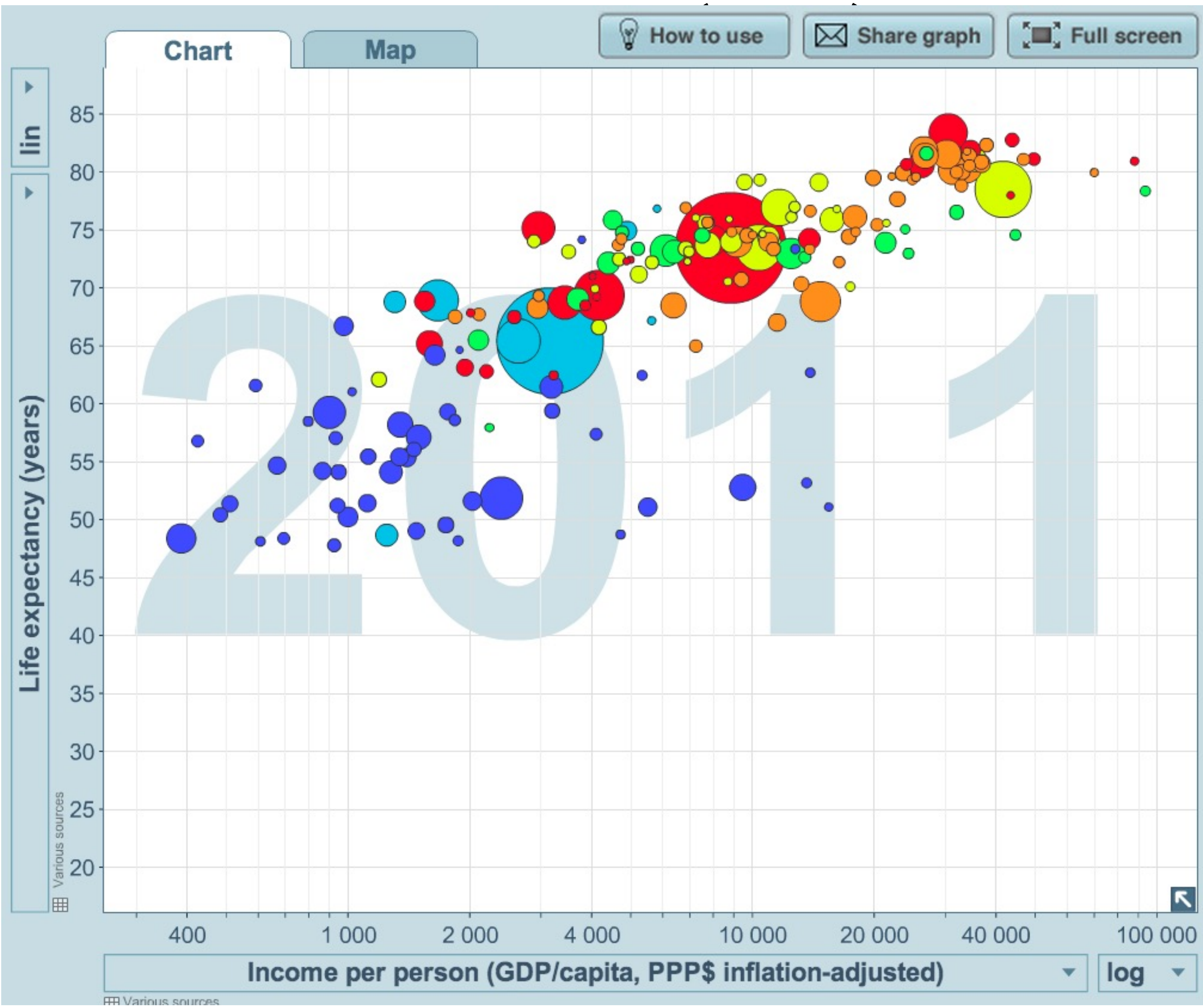
Transformations

- Linear transformations
  - Predicting variables
  - Response variables
- Log transform
  - Logarithms
  - Log transforming response variables
  - Log transforming predicting variables
  - Log transforming response and predicting variables
- Logit transform (maybe today, maybe later in logistic)
  - Logit and logistic transformations (inverses of each other)
  - Logit(y) ~ x
  - Y~logit(x) ?

Transforming predictor variables...



Transforming predictor variables...



When to log transform *predictor* variables?

- When proportional changes in x yield constant changes in y.
  - E.g., income/wealth
- When x is very positively skewed
- When x is bounded at 0 and is close to it

- These tend to co-occur.

Log-transforming predictor variable

$$Y_i = \beta_0 + \beta_1 \log_{10}(X_{1i}) + \varepsilon_i$$

- So what does a slope of $B_1 = 2$ mean?
  - For every unit increase of log10(X1) (all else equal), Y increments by 2.
  - For every increase of X1 by a **factor of 10** (all else equal) Y increments by 2.

Transformations

- Linear transformations
  - Predicting variables
  - Response variables
- Log transform
  - Logarithms
  - Log transforming response variables
  - Log transforming predicting variables
  - Log transforming response and predicting variables
- Logit transform (maybe today, maybe later in logistic)
  - Logit and logistic transformations (inverses of each other)
  - Logit(y) ~ x
  - Y~logit(x) ?

Log transforming response and predictor

- When proportional changes in x yield proportional changes in y.
  - E.g., doubling x causes quadrupling in y
  - a power law relationship
    y ~ b*x^a
    log(y) ~ a*log(x)+log(b)

  - Interpretation of slope / intercept:
    - Slope: exponent of power law relationship
      1: y proportional to x. 2: y proportional to x^2, etc.
    - Intercept: proportionality constant
      y = intercept*(x^slope)

Power Law: Kleiber's law

metabolic.rate ~ mass^(3/4) + c

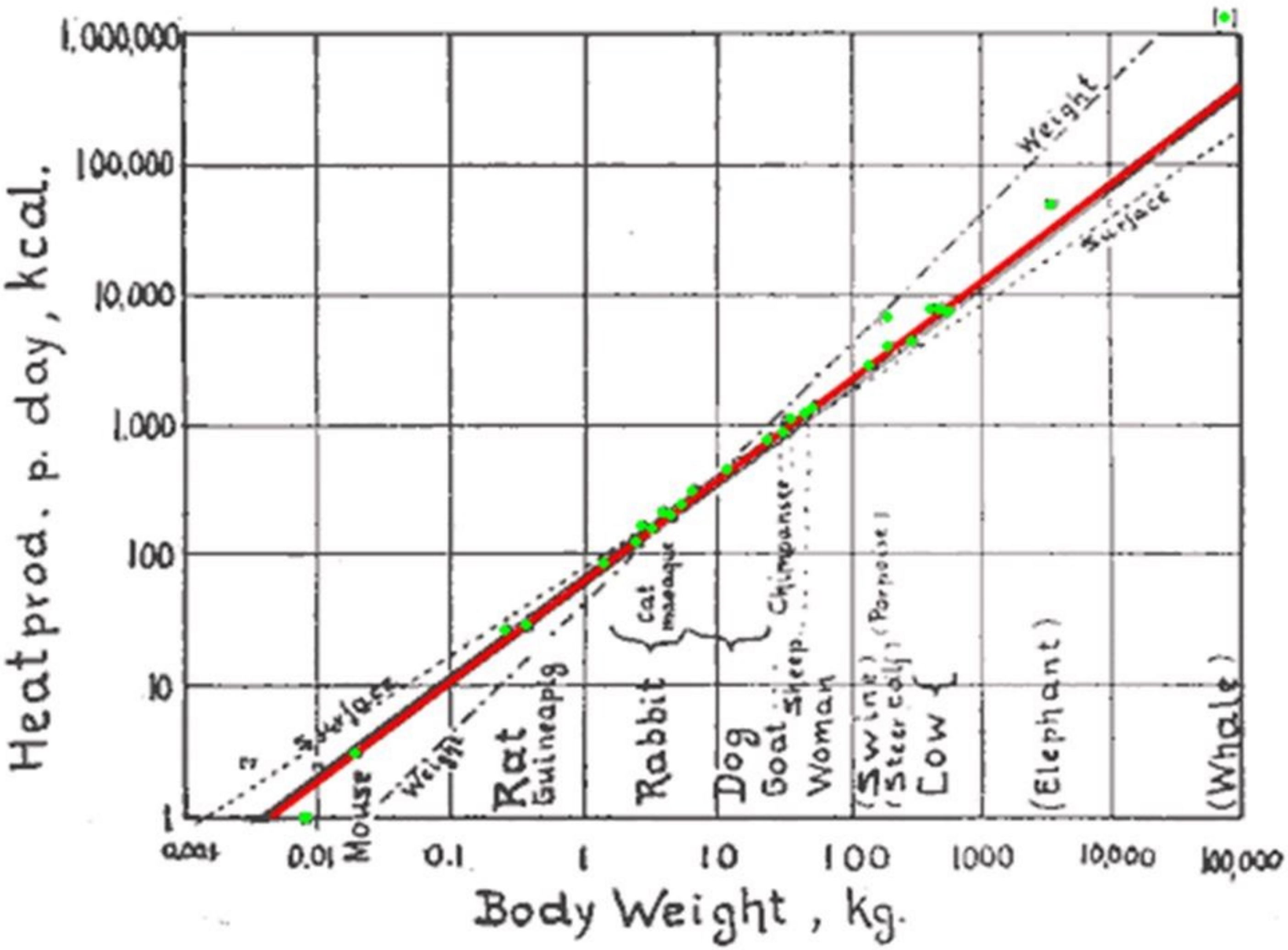


Fig. 1. Log. metabol. rate/log body weight

Log-linearized regression.

For each of these: how would you set up the regression, what would you expect the coefficients to be, what do they mean, and what do you expect the R^2 to be?

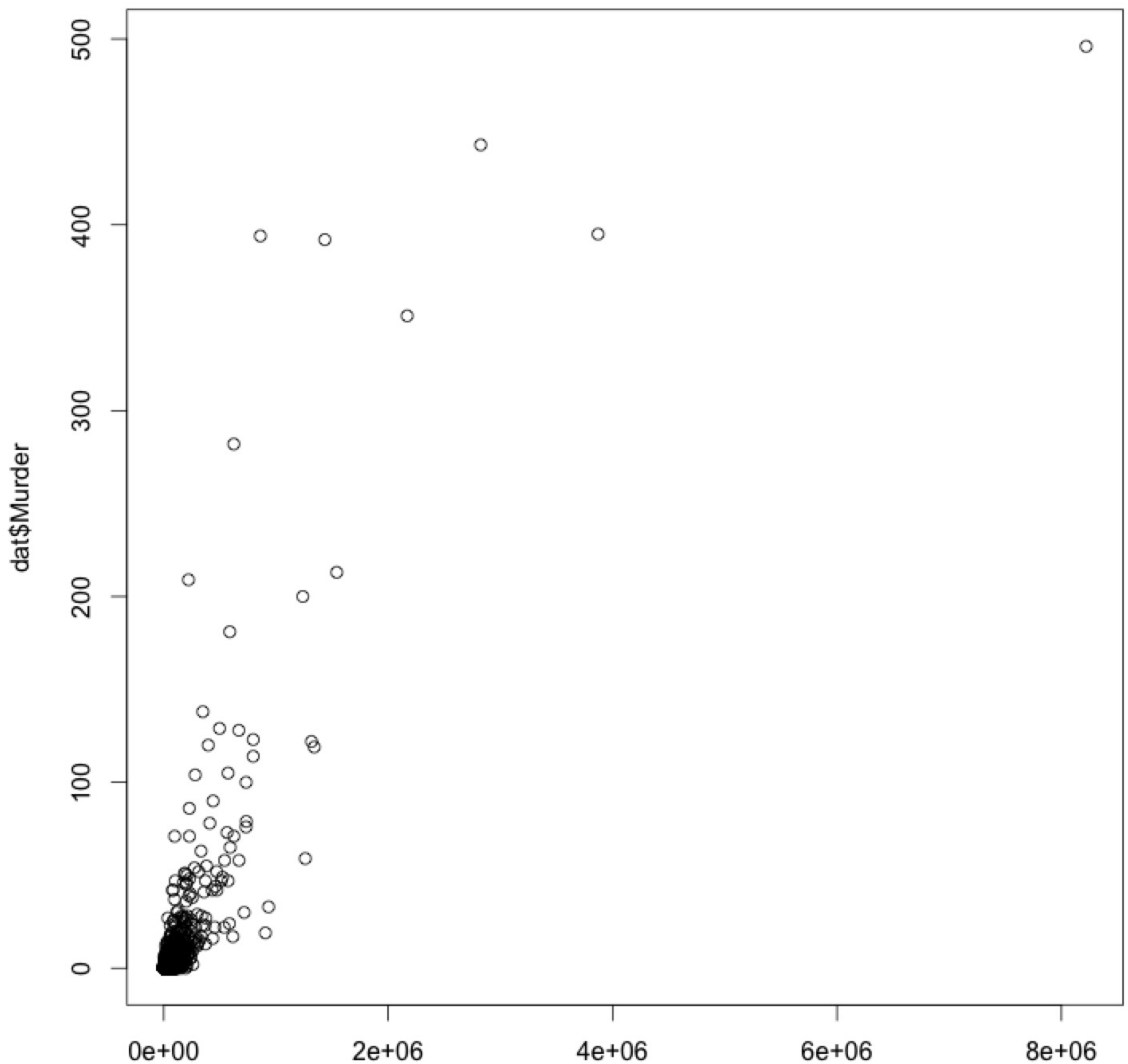
1) We are predicting number of murders as a function of city/town population size.
2) We are predicting theft rate (crimes per 100,000) as a function of the church density (churches per 100,000).
3) We are predicting time to solve a math problem as a function of GRE score.
4) We are predicting human weight as a function of human height.
5) We are predicting iq as a function of cranial volume.

Log-linearized regression.

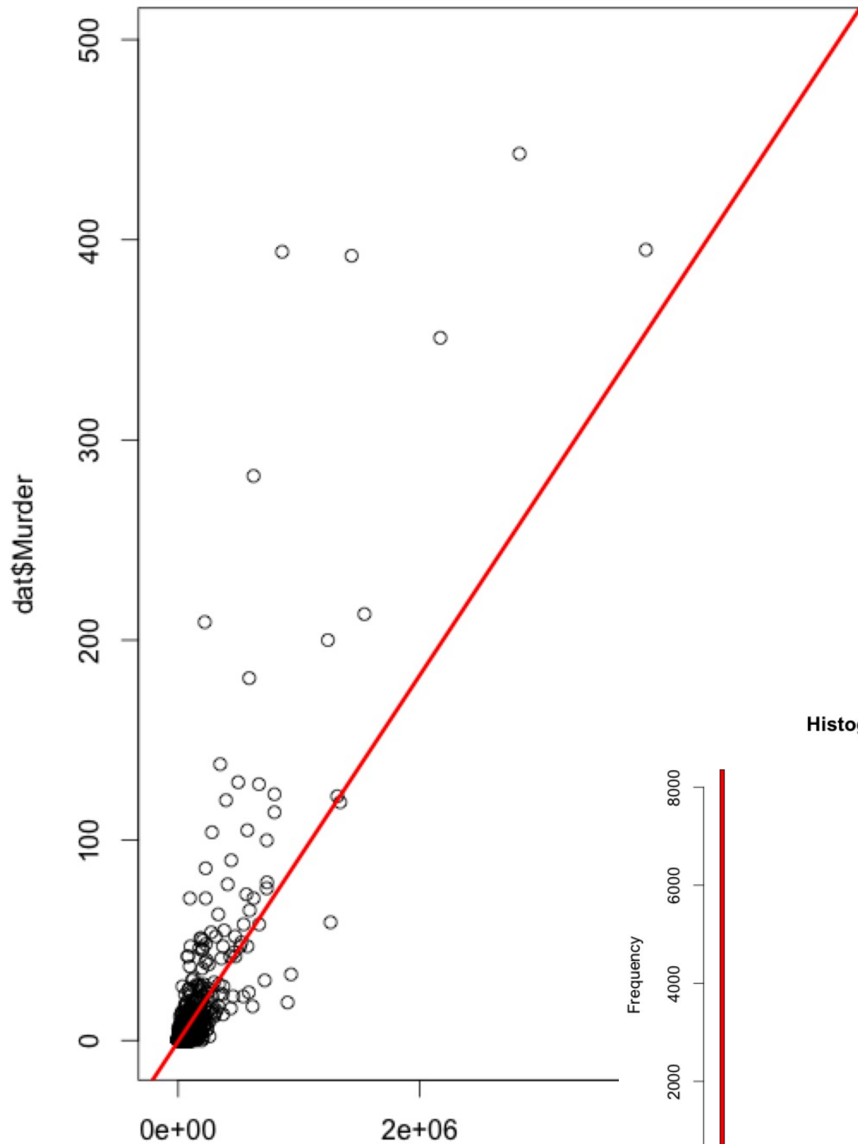
For each of these: how would you set up the regression, what would you expect the coefficients to be, what do they mean, and what do you expect the R^2 to be?

1) We are predicting number of murders as a function of city/town population size.

Murders vs City Population



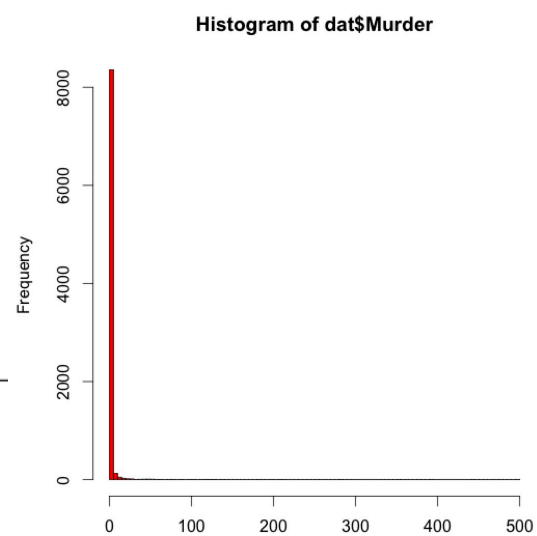
Murders vs City Population

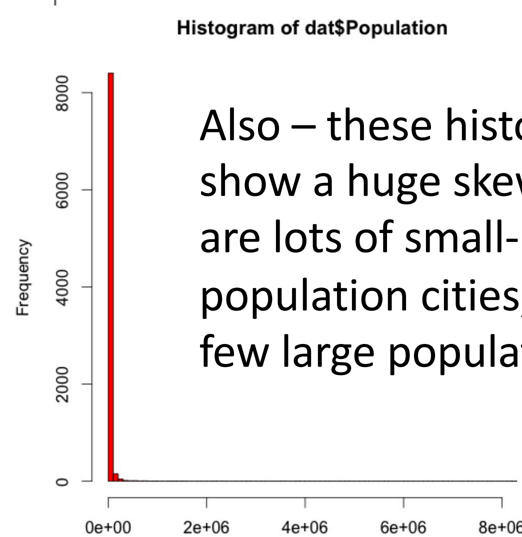



```
Call:
lm(formula = dat$Murder ~ dat$Population)

Coefficients:
   (Intercept)  dat$Population
    -5.033e-01       9.149e-05
```

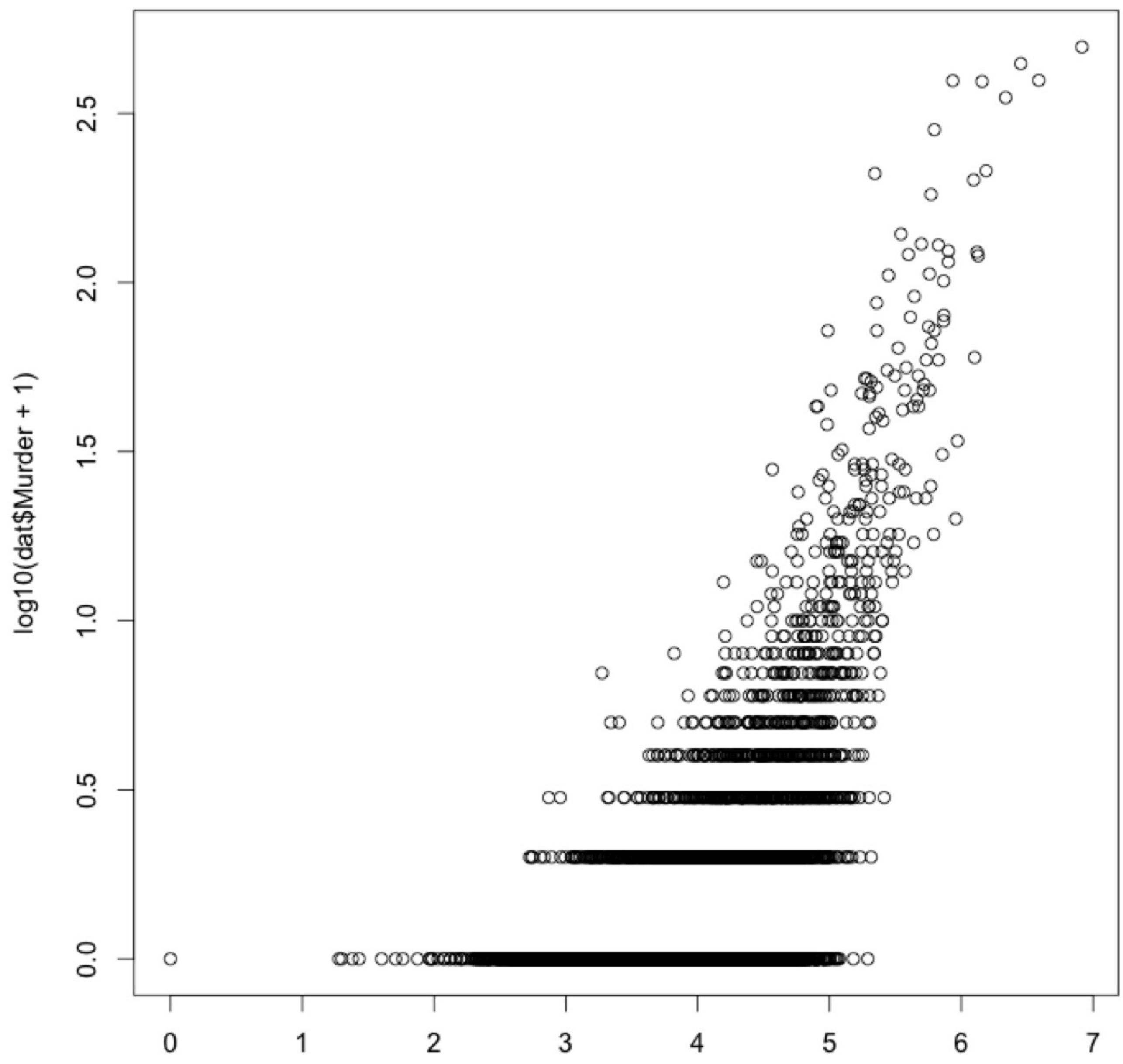
That looks a bit off…
Why? Because treating population and murder counts as linear makes large (outlier?) values have too much of an effect.

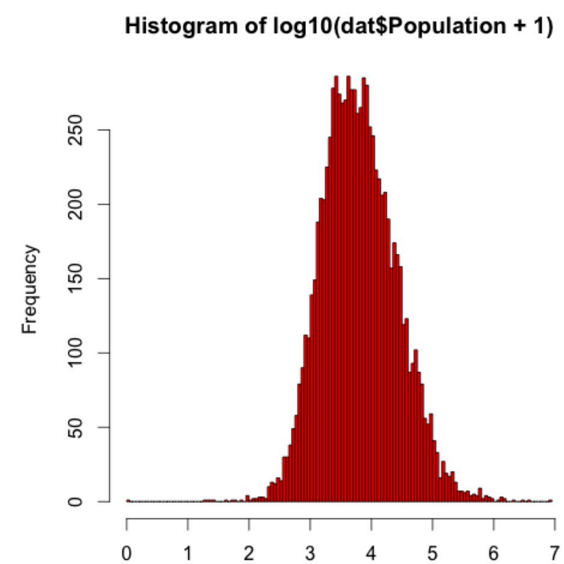




Also – these histograms show a huge skew: there are lots of small-population cities, and very few large population cities.

Log(# Murders) vs log(City Population)





Transformations

- Linear transformations
  - Predicting variables
  - Response variables
- Log transform
  - Logarithms
  - Log transforming response variables
  - Log transforming predicting variables
  - Log transforming response and predicting variables
- Logit transform (maybe today, maybe later in logistic)
  - Logit and logistic transformations (inverses of each other)
  - Logit(y) ~ x
  - Y~logit(x) ?

Nonlinear Transformations

Log transform response variable:
$\text{Log}(y) \sim b_0 + b_1 x_1 + \ldots$

- Because…
  - …predictors make proportional changes to y
  - …y has a large positive skew
  - …y is bounded at (and close to) 0
  - …y covers many orders of magnitude
- Suggestions: Use base 10 log.
- Consequences: exponential relationship
  slopes now mean: per unit increase in x…
  - …$\log_{10}(y)$ goes up by a constant B1
  - …y goes up by a factor of 10^B1 (exp(B1) if ln)

Nonlinear Transformations

Log transform predictor variable:
$y \sim b_0 + b_1 * \log(x_1) + \ldots$

- Because…
  - …response y is sensitive to proportional changes of x
  - …x has a large positive skew
  - …x is bounded at (and close to) 0
  - …x covers many orders of magnitude
- Suggestions: Use base 10 log.
- Consequences: logarithmic relationship slopes: constant B_1 increment to Y for every…
  - unit increment in $\log_{10}(x)$, or…
  - x10 multiplication (proportional change) of x

Nonlinear Transformations

Log transform response and predictor:
log(y) ~ b0 + b1*log(x1) + …

- Because…
  - …proportional changes to x yield proportional changes to y
  - …x and y have positive skew
  - …x and y are bounded at (and close to) 0
  - …x and y cover many orders of magnitude
- Suggestions: Use base 10 log.
- Consequences: power law relationship
  $\log_{10}(y) = (\text{intercept}) + (\text{slope})*\log_{10}(x)$
  y=10^(intercept) * x^slope

Nonlinear Transformations

- Log transform response variable:
  $\log(y) \sim b_0 + b_1 * x_1 + \ldots$ **Adding to X -> Multiplying Y**
- Log transform predictor variable:
  $y \sim b_0 + b_1 * \log(x_1) + \ldots$ **Multiplying X -> Adding to Y**
- Log transform response and predictor:
  $\log(y) \sim b_0 + b_1 * \log(x_1) + \ldots$ **Multiplying X -> Multiplying Y**
- Logit transform response variable:
  $\text{logit}(y) \sim b_0 + b_1 * x_1 + \ldots$
- Logit transform predictor variable:
  $y \sim b_0 + b_1 * \text{logit}(x_1) + \ldots$
- Logit transform response and predictor:
  $\text{logit}(y) \sim b_0 + b_1 * \text{logit}(x_1) + \ldots$

Log-linearized regression.

Interpret the coefficients/predictions for these regressions

1) life.expectancy ~ log10(GDP/capita)*9+35
2) log10(city.GDP/capita) ~ -0.4*corruption.index + 0.5*$\log_{10}$(population/mi^2) + 2.5
   [corruption.index = {-5 to 5} survey corruption prevalence estimate]
3) $\log_{10}$(voter.turnout) ~ 0.5*$\log_{10}$(population) + 0.8*pres+ 0.2*sen + 0.4*gov - 1
   [pres = {1,0} whether it is a presidential election]
   [sen = {1,0} whether it is a senate election]
   [gov = {1,0} whether it is a state governor election]
4) log10(RT) ~ accuracy – age{y}
   [accuracy = {1,0}]
5) adult.IQ ~ -5*weeks.premature + 8*breast.fed + 4*$\log_{10}$(mean.daily.calories) + 93
   [breast.fed = {1,0} whether was breast fed as infant]

Transformations

- Linear transformations: don't change the regression, but make the coefficients more user-friendly.
- Log transformations: Linearize variables to make a linear regression behave like an exponential, logarithmic, or power law relationship. (proportional changes matter for logs)
- Variable-combination transformations: help extract more useful variables from ones that are perhaps correlated, or susceptible to extraneous fluctuations.
- In practice,
  - all of these can (and should) be used in combination, but not in a fishing expedition: in a thoughtful theoretical manner.
  - Check scatter-plots and histograms, to look for desirable transformations.

Wage gap data (2013 BLS)

```
bls <- read_csv('http://vulstats.ucsd.edu/data/BLS.2016.csv')
```

For each *Occupation* it shows the occupation *Category,* how many people have this occupation **.n (in 1000s),* median weekly earnings **.earn*, and std. err of earnings **.earn.se.* for everyone (*all.**), females (*f.**), and males (*m.**).

Characterize as best as you can the relationship between male and female median weekly wages.

Consider:

- If you were to come up with just one number, of the form "women make x% of what men make", how would you do it?
- What kinds of relationships can you capture regressing female~male wages with different transforms? Which formulation makes more sense a priori?
- What does the slope mean?
- What does the intercept mean? Should it be free to vary? What happens if you fix it?

Transformations

- Linear transformations
  - Predicting variables
  - Response variables
- Log transform
  - Logarithms
  - Log transforming response variables
  - Log transforming predicting variables
  - Log transforming response and predicting variables
- Logit transform (maybe today, maybe later in logistic)
  - Logit and logistic transformations (inverses of each other)
  - Logit(y) ~ x
  - Y~logit(x) ?

Why do a logit transform?

- If variable is bounded between 0 and 1…
  or between any two values, and then rescaled to [0 1]; Most often: proportions (accuracy, etc.)
  - A linear model will not work well – doesn't respect bounds.
  - It usually gets progressively 'harder' to get closer and closer to the bound
    0.98 to 0.99 is a bigger 'change' than 0.58 to 0.59
    e.g., improving from 50th to 55th percentile is relatively easy, from 90th to 95th is much harder (in anything!)
- Logit (or, log-odds) transform fixes both problems.
  - Transforms variables from [0 1] to [-infinity +infinity]
    so now a linear model works fine.
  - Log-odds differences for identical proportion increments are bigger near the bounds:
    (0.50 -> 0.51): +0.04 log odds; (0.90 -> 0.91): +0.12 log odds

Odds

- P : proportion or probability of outcome
  - E.g. P(male), P(correct), etc.
- P/(1-P) : Odds of outcome
  - Probability of getting outcome divided by probability of not getting outcome
  - To go back to probability from odds: P = Odds / (1+Odds)
    Odds = 4; P = 0.8
    Odds = 1/5; P = 1/6

Log Odds

- Odds is on scale [0 Infinity] – a ratio
- Log transforming linearizes (to scale [-inf inf])
- This is usually done with natural (base e) log. Let's keep it that way.

- Log[odds] = 3
  Log[p/(1-p)] = 3
  p/(1-p) = odds = exp(3) = 20
  p = 20/21 = 0.95

Probability, Odds, Log-odds

- P : proportion or probability of outcome
  - Possible values: [0 1]

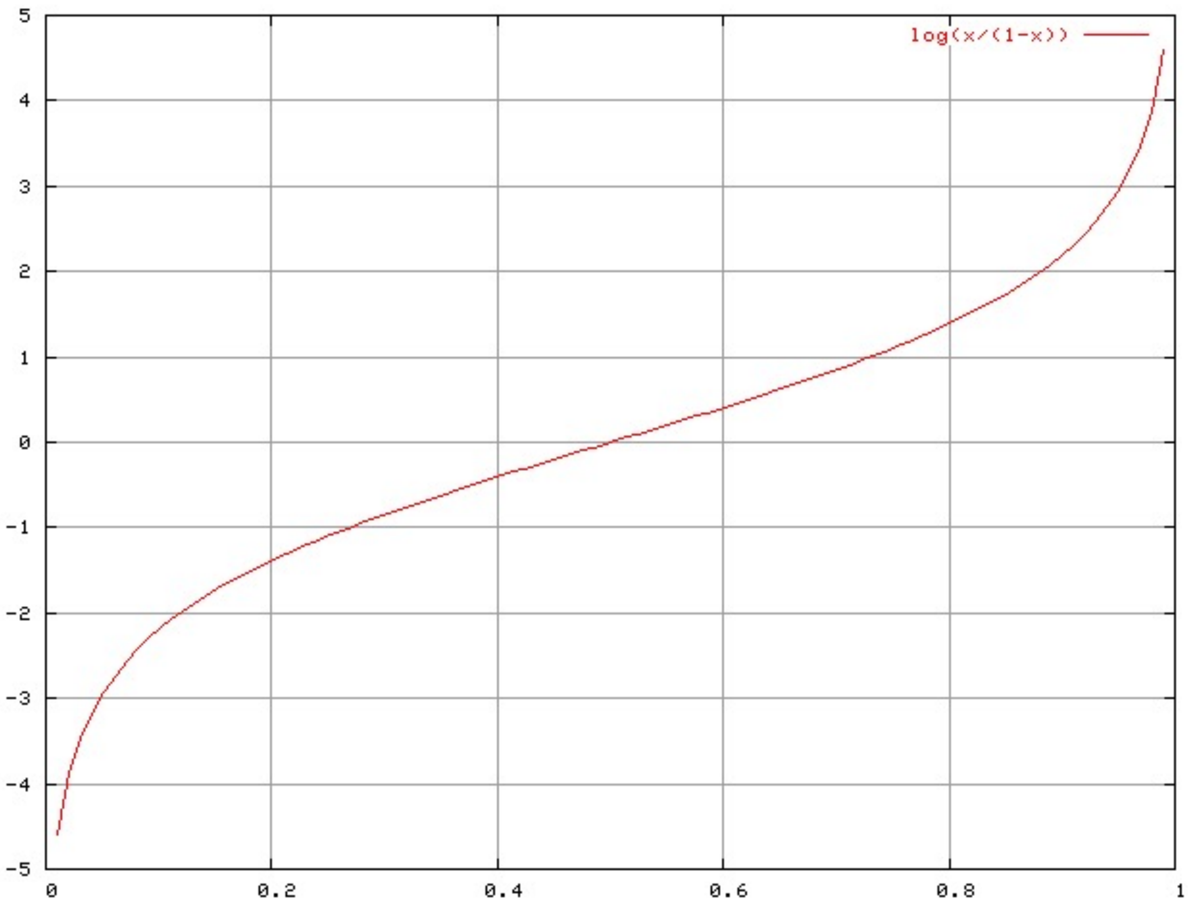
- P/(1-P) : Odds of outcome
  - Possible values: [0 +infinity]
  - To go back to probability from odds: P = Odds / (1+Odds)

- Log[P/(1-p)] : Log-odds of outcome
  - Possible values: [-infinity +infinity]
  - To go back to odds from log-odds: Odds = exp[log.odds]

Logit (log odds) transform

- From [0 1] -> [-inf inf]

$$\text{logit}(p) = \log\left(\frac{p}{1-p}\right) = \log(p) - \log(1-p).$$



Logit transforms

Logit transform of response variables logit(y)~...

- Useful when modeling proportions and changes in proportions – after logit transform you can use linear model to describe changes in logit(p)
  - This is the basis of logistic regression (later)

Logit transform of predictor variables y~logit(x)...

- (sometimes) useful when a proportion is a predictor
  - not very often done:
  - Not necessary since in linear model we don't much care about bounds on x.
  - Sometimes the non-linear difficulty of being closer to bounds makes this transform better account for the data.

Logit transform of response variable

$$\log\left(\frac{y}{1-y}\right) = B_0 + B_1 x_1 + B_2 x_2$$

- What do the coefficients mean?

Logit transform of response variable

$$\log\left(\frac{y}{1-y}\right) = B_0 + B_1 x_1 + B_2 x_2$$

- What do the coefficients mean?
  - Interpretation via log odds:
  - B1 means: per unit of x1, log odds increment by B1
  - B0 means: log odds of outcome when all x=0

Logit transform of response variable

$$\log\left(\frac{y}{1-y}\right) = B_0 + B_1 x_1 + B_2 x_2$$

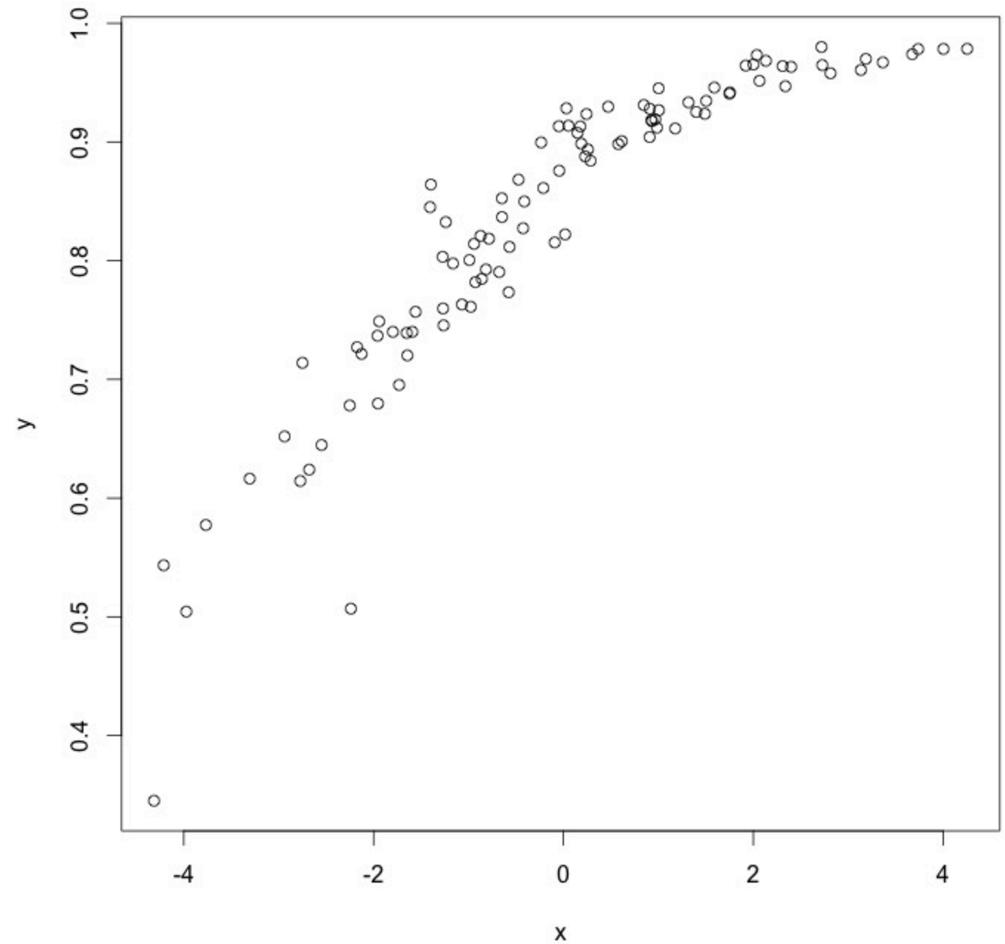
- What do the coefficients mean?
  - Interpretation via log odds:
    - B1 means: per unit of x1, log odds increment by B1
    - Bo means: log odds of outcome when all x=o

$$\left(\frac{y}{1-y}\right) = \exp(B_0) * \exp(B_1) \wedge x_1 * \exp(B_2) \wedge x_2$$

  - Interpretation via odds:
    - B1 means: per unit increment of x1, odds **multiply** by exp(B1)
    - Bo means: when all x=o odds of outcome are exp(Bo)

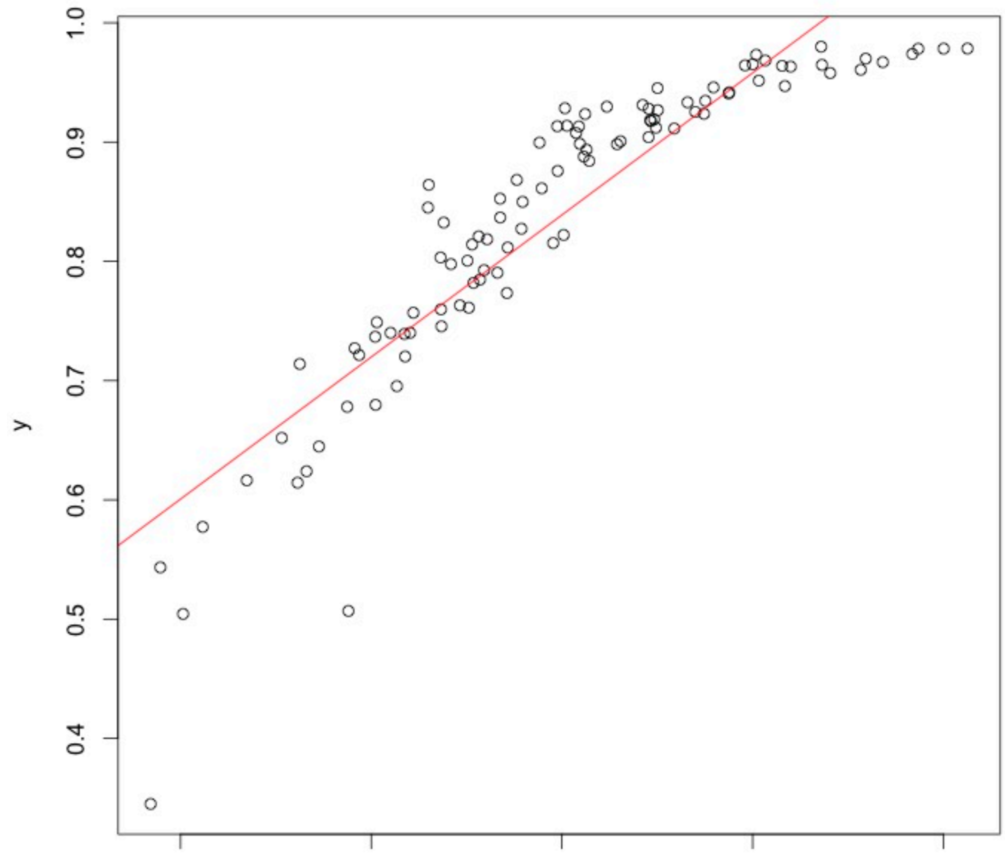
Logit regression

Y is some proportion
(e.g., GRE percentile)
and x is some predictor
(e.g., study time)



Logit regression

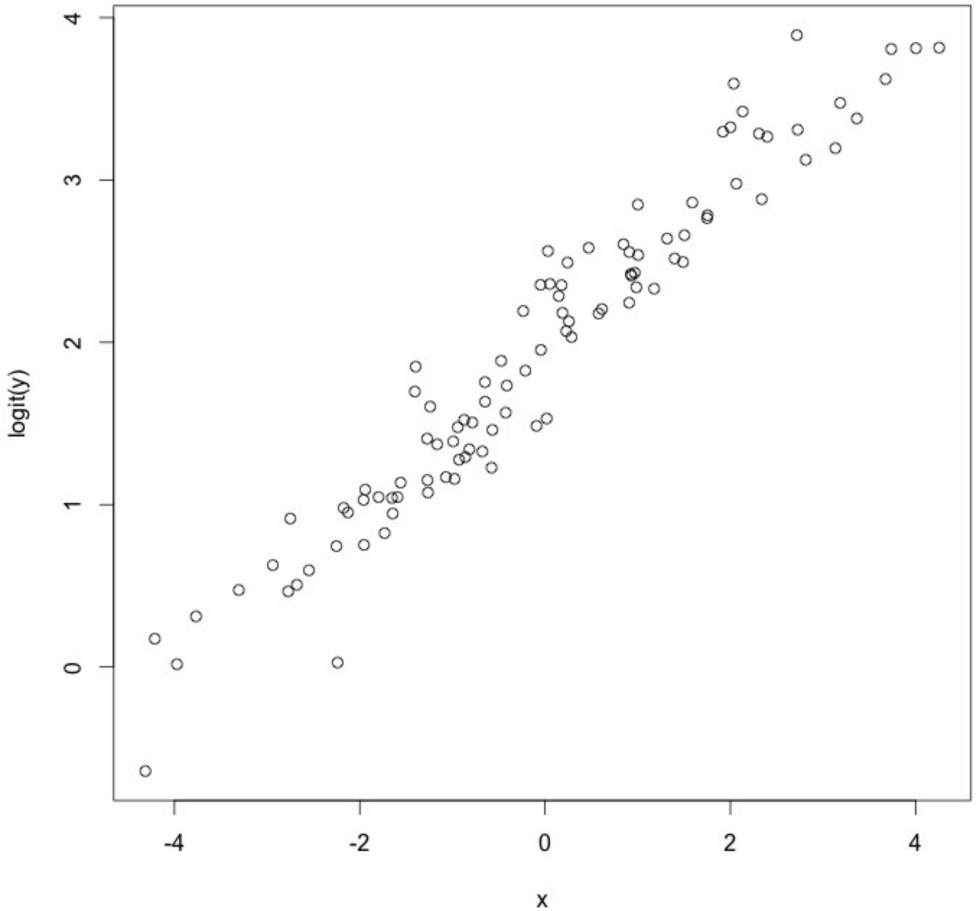
Y is some proportion
(e.g., GRE percentile)
and x is some predictor
(e.g., study time)




```
summary(lm(y~x))
```

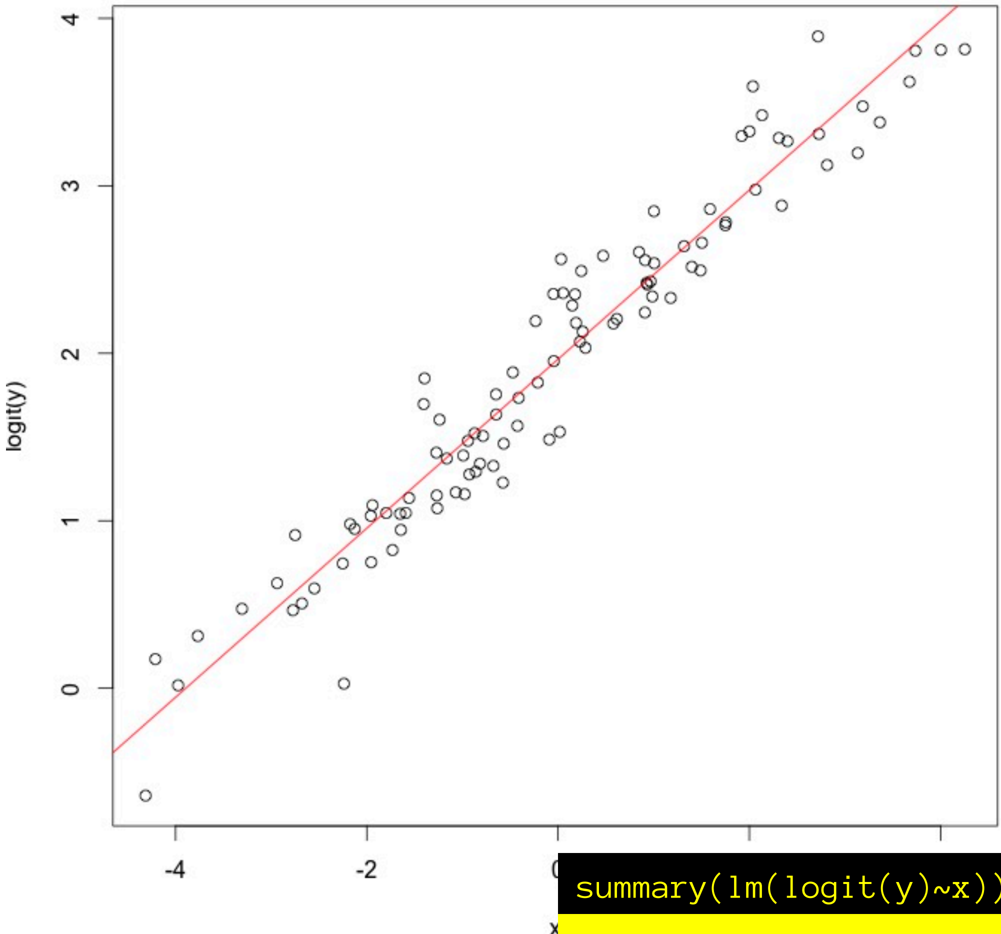
```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 0.839894   0.005363  156.61   <2e-16 ***
x           0.059567   0.002809   21.20   <2e-16 ***
---
Residual standard error: 0.05361 on 98 degrees of freedom
Multiple R-squared: 0.821,     Adjusted R-squared: 0.8192
F-statistic: 449.6 on 1 and 98 DF,  p-value: < 2.2e-16
```

Logit regression




```
logit = function(p){log(p/(1-p))}
plot(logit(y),x)
```

Logit regression




```
logit = function(p){log(p/(1-p))}
plot(logit(y),x)
```

```
summary(lm(logit(y)~x))
```

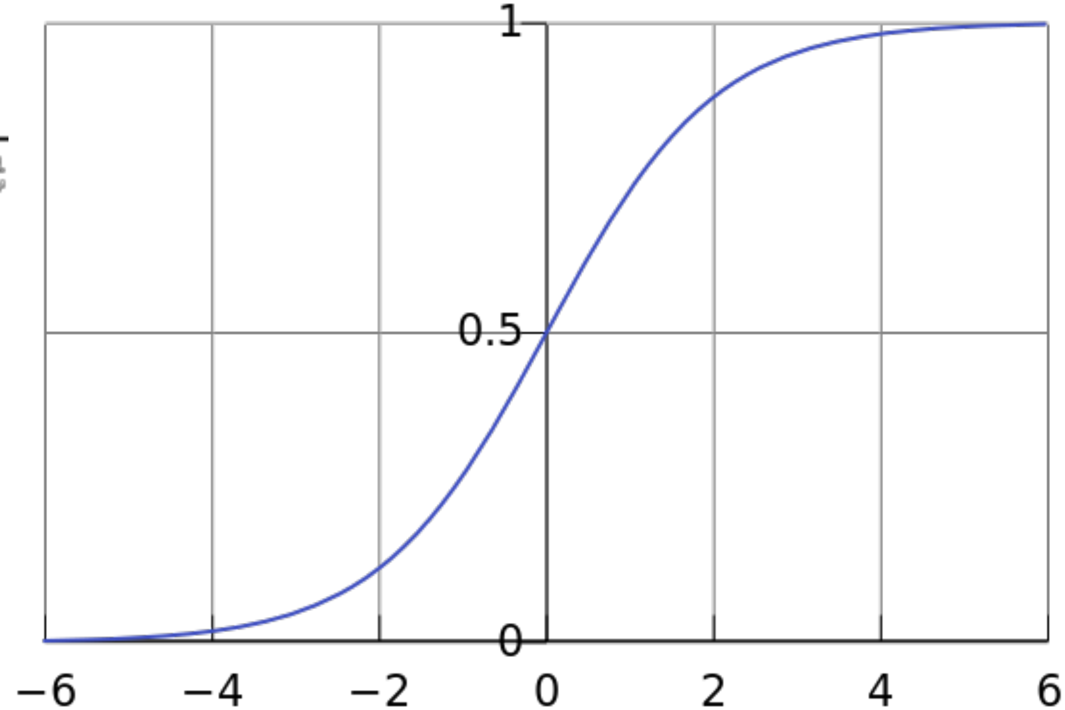
```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.96476    0.02582   76.08   <2e-16 ***
x            0.50464    0.01353   37.30   <2e-16 ***

Residual standard error: 0.2581 on 98 degrees of freedom
Multiple R-squared: 0.9342,    Adjusted R-squared: 0.9335
F-statistic:  1392 on 1 and 98 DF,  p-value: < 2.2e-16
```

Logistic transform: undoing the logit

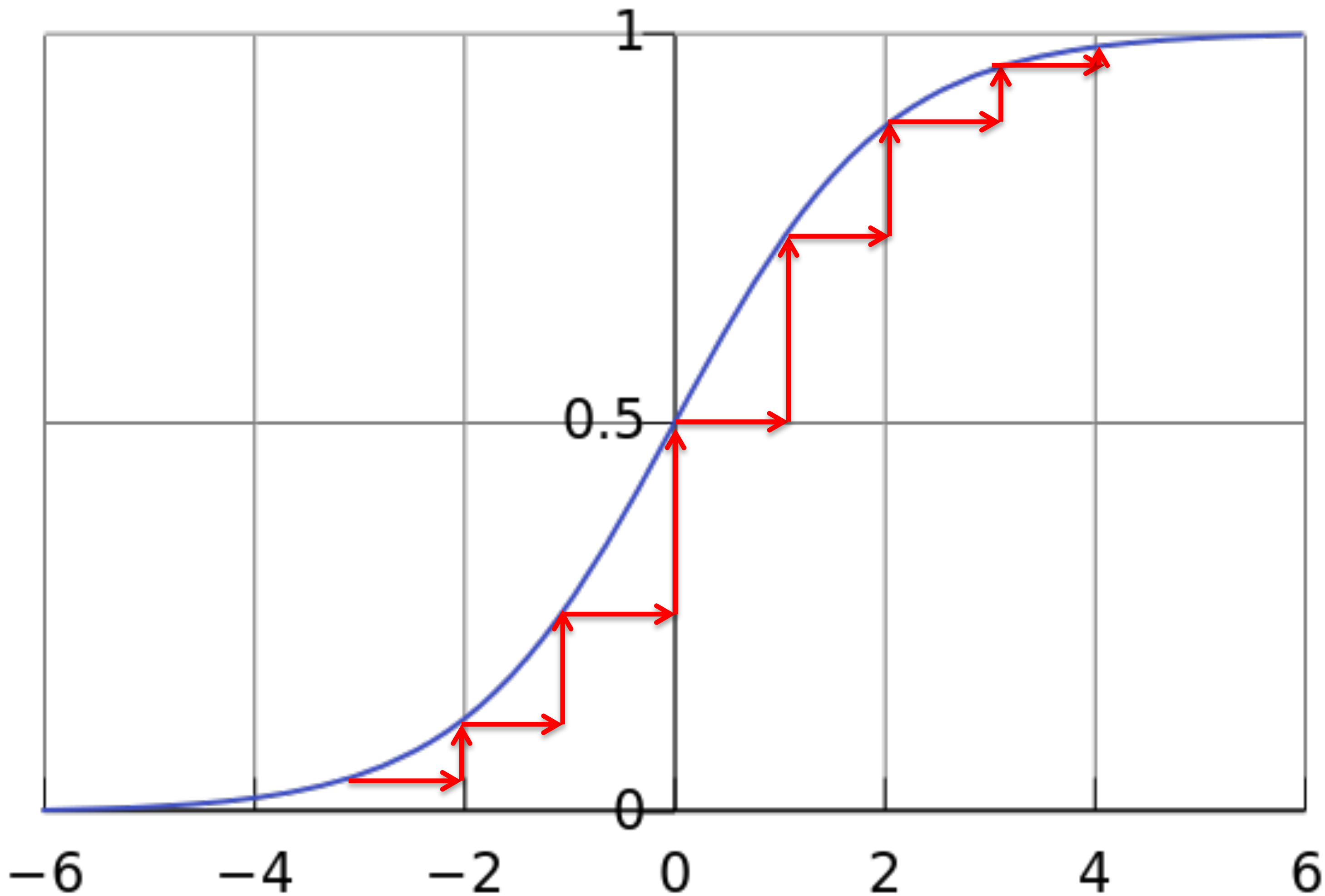
- Logit(p)=x   p in [0 1]   -> x in [-inf inf]
- Logistic(x)=p   x in [-inf inf]   -> p in [0 1]
- Logistic(Logit(p)) = p
- Logit(Logistic(x)) = x

$$P(t) = \frac{e^t}{e^t + 1} = \frac{1}{1 + e^{-t}}$$

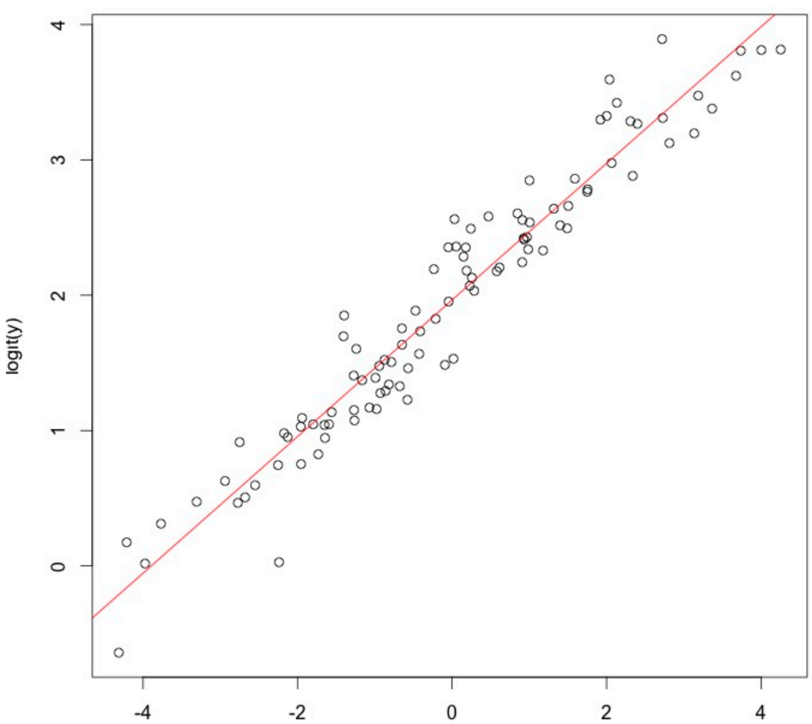


Logistic transform: undoing the logit

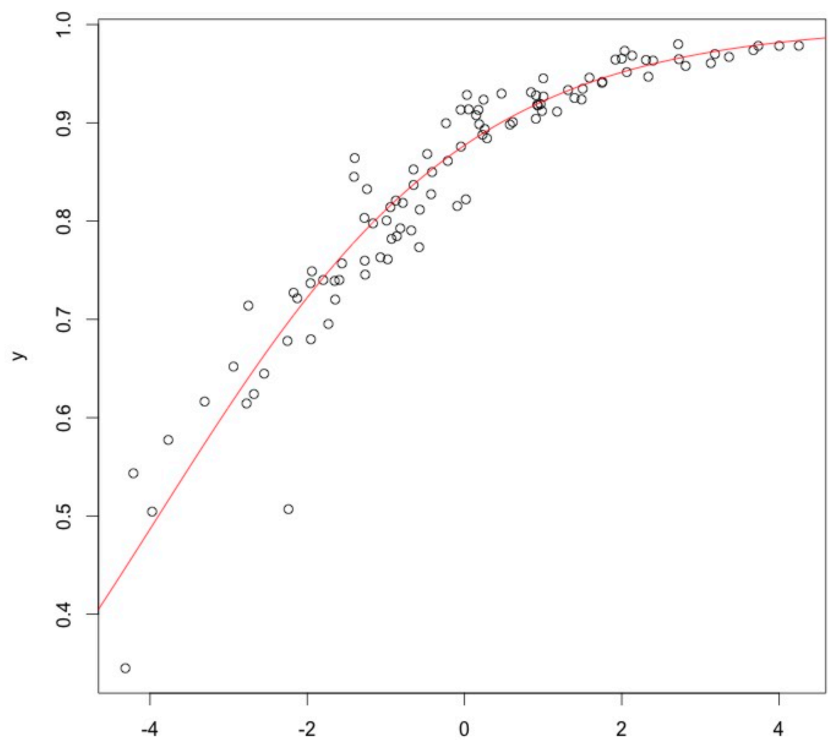
- Logit(y)~B1*X+Bo
- Slope: increment 1 unit on logistic (log[odds]) scale



Logit regression in probability space



Straight line in logit (log-odds) units yields a curved (sigmoidal) line in probability




```
logit = function(p){log(p/(1-p))}
logistic = function(z){1/(1+exp(-z))}
```

```
Pred.log.odds = x*B1+B0
Pred.probability = logistic(Pred.log.odds)
```

"Smoothing"

If y = 0 or 1, logit(y) is undefined:

y/(1-y) = 0 or infinity; so log(y/(1-y)) is undefined/infinite.

What to do in this case?

- Option 1: Give up on logit.
  - Not ideal if other reasons favor logit.
- Option 2: Smooth by adding constant to p and 1-p: y' = (y+e)/(1+2e)
  - How to choose e?
    - In case of empirically calculated proportions, easy to postulate two unobserved points: (one success and one failure). This brings all proportions a bit closer to 0.5 for accuracy: y = correct/total. y' = (correct+1)/(total+2)
    - Otherwise, make e small (e.g., smaller than smallest y or (1-y)).

Working with logit regression

When: y is a proportion (or is bounded and scaled)

Why: because we assume that changes in log-odds are linear with our predictors.

not unreasonable, may not be exactly right, but the alternative (that proportion is linear in our predictors) is definitely wrong

How:

$$\text{logit}(y) \sim B_0 + B_1 x_1 + B_2 x_2 + \ldots + \text{error}$$

Alternatively (but not practically in lm()):

$$y \sim \text{logistic}(B_0 + B_1 x_1 + B_2 x_2 + \ldots + \text{error})$$

Cautions: Coefficients are tricky. Per unit increment in x...

- Log-odds(y) [logit(y)] increments by a constant B_1
- Odds(y) multiplies by a factor of $\exp(B_1)$
- y has no constant change (because proportional odds changes have different impacts on y depending on its initial value)

Nonlinear Transformations

- Logit transform predictor variable:
  $y \sim b_0 + b_1 * \text{logit}(x_1) + \ldots$
- Because…
  - …x is a proportion or is bounded (and scaled to [0 1] range)
  - …y should change linearly with log-odds of x.
- …rarely used!

Nonlinear Transformations

- Logit transform response and predictor:
  logit(y) ~ b0 + b1*logit(x1) + ...
- Because...
  - Log-odds of x and log-odds of y are linearly related...
- ...rarely used!

Practice

1) Our regression predicts that logit(GRE percentile) will be 1.6, what is the GRE percentile?

In a regression predicting logit(proportion correct) from IQ, we find a slope of 0.5, and an intercept of -50....

2) What will be the proportion correct for someone with an IQ of 80?

3) How will accuracy change when increasing IQ by 10 points?

4) What will be the difference in accuracy between those with an IQ of 100 and those with an IQ of 110?

5) What will be the difference in accuracy between those with an IQ of 140, and those with an IQ of 150?

6) Is the test that we are using here useful for assessing IQ? In what range?

Nonlinear Transformations

- Log transform response variable:
  $Log(y) \sim b_0 + b_1 x_1 + \ldots$
- Log transform predictor variable:
  $y \sim b_0 + b_1 * log(x_1) + \ldots$
- Log transform response and predictor:
  $log(y) \sim b_0 + b_1 * log(x_1) + \ldots$
- Logit transform response variable:
  $logit(y) \sim b_0 + b_1 * x_1 + \ldots$
- Logit transform predictor variable:
  $y \sim b_0 + b_1 * logit(x_1) + \ldots$
- Logit transform response and predictor:
  $logit(y) \sim b_0 + b_1 * logit(x_1) + \ldots$

These are sometimes called "linearized" regression, because we can capture a non-linear relationship using the linear model by using a non-linear transformation.

Transformations

- Linear transformations
  – Predicting variables
  – Response variables
- Log transform
  – Logarithms
  – Log transforming response variables
  – Log transforming predicting variables
  – Log transforming response and predicting variables
- Logit transform (maybe today, maybe later in logistic)
  – Logit and logistic transformations (inverses of each other)
  – Logit(y) ~ x
  – Y~logit(x) or logit(y)~logit(x)?