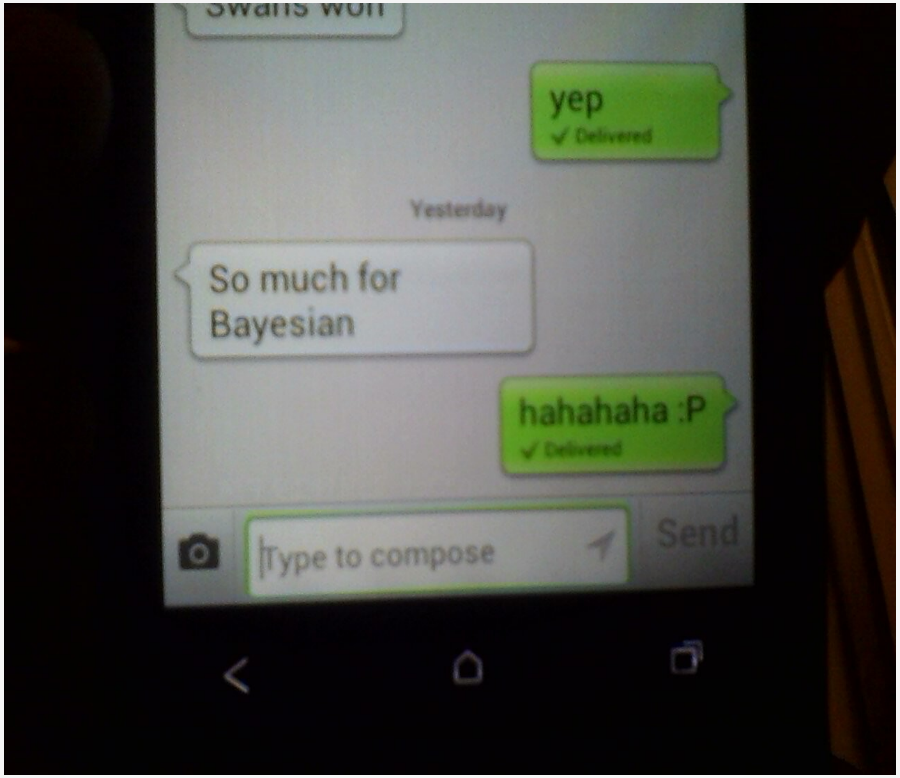
STATS 331

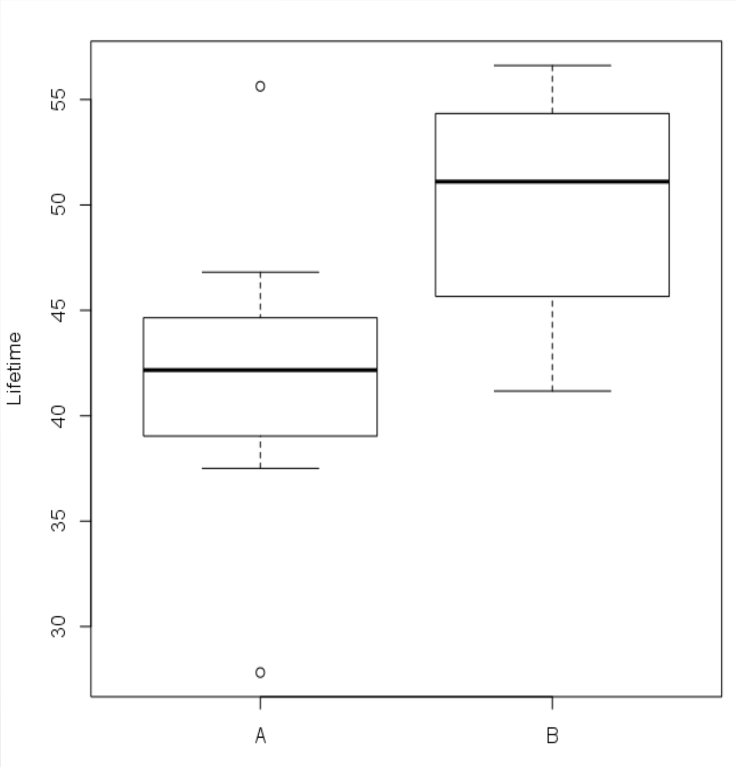


Introduction to Bayesian Statistics
Semester 2, 2016

Today's Lecture

- “One way ANOVA” model – a generalisation of the “t-test” models to more groups

- Posterior predictive checks



Reminder of t-test model 3 – Likelihood

```
for(i in 1:N1)
{
  x1[i] ~ dnorm(mu1, 1/sigma^2)
}
for(i in 1:N2)
{
  x2[i] ~ dnorm(mu2, 1/sigma^2)
}
```

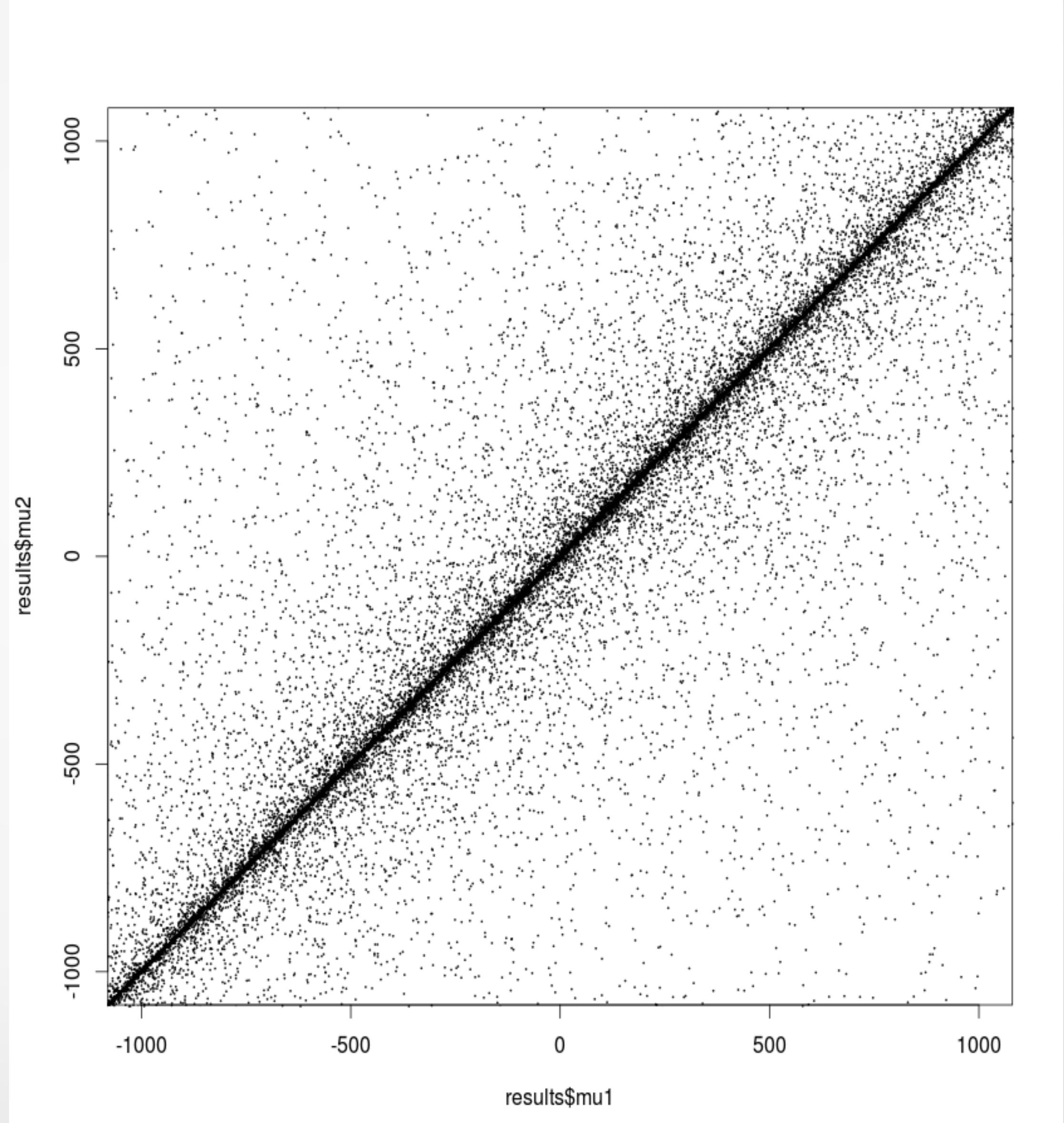
Reminde of t-test Model 3 – Priors

```
# Priors for the parameters
mu1 ~ dnorm(grand_mean, 1/diversity^2)
mu2 ~ dnorm(grand_mean, 1/diversity^2)

# Priors for the hyperparameters
grand_mean ~ dnorm(0, 1/1000^2)
log_diversity ~ dunif(-10, 10)
diversity <- exp(log_diversity)
```

Reminder of t-test model 3

"$\boldsymbol{\mu_1}$ and $\boldsymbol{\mu_2}$ aren't precisely equal, that's silly. But they could be very similar. Or not."



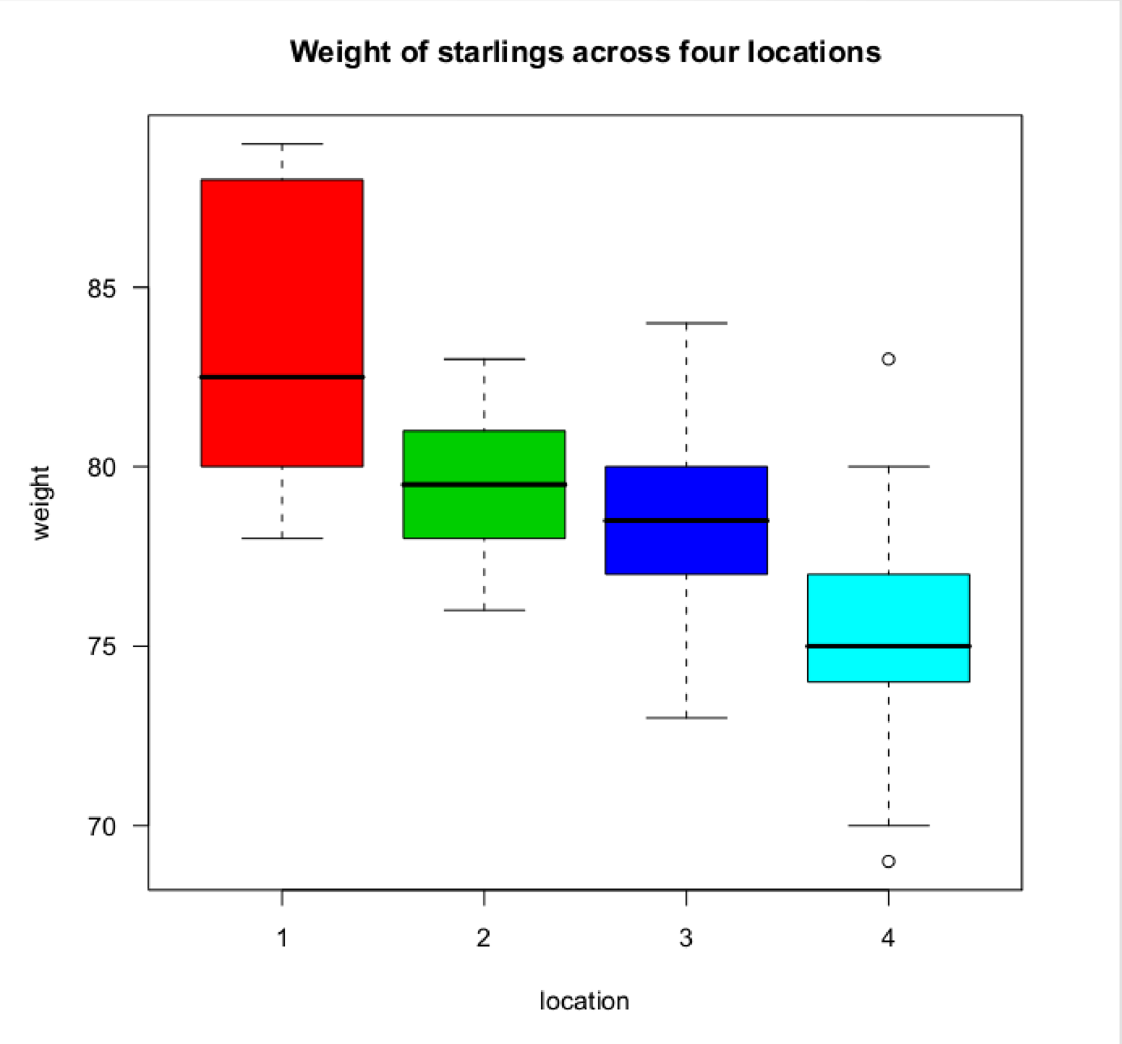
One-Way ANOVA

ANOVA stands for “Analysis of Variance”, but we won't be analysing any variance.

My secret confession

Example

Weights of starlings at four locations



Locations don't have any particular **order**

Reformat Data

Our life will be easier if we put all the data into one vector, instead of having a separate vector for each group.

```
data = list(x = c(78, 88, 87, 88, 83, 82,
  81, 80, 80, 89, 78, 78, 83, 81, 78, 81,
  81, 82, 76, 76, 79, 73, 79, 75, 77, 78,
  80, 78, 83, 84, 77, 69, 75, 70, 74, 83,
  80, 75, 76, 75), group = c(1, 1, 1, 1, 1,
  1, 1, 1, 1, 1, 2, 2, 2, 2, 2, 2, 2, 2, 2,
  2, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 4, 4, 4,
  4, 4, 4, 4, 4, 4, 4), N = 40)
```

Likelihood

```
# Likelihood
for(i in 1:N)
{
    # Pick out the appropriate mu parameter
    # for this data point
    theta[i] <- mu[group[i]]
    x[i] ~ dnorm(theta[i], 1/sigma^2)
}
```

Likelihood

- This sure beats having to write 4 loops in the JAGS model!

- Note: the `mu`s will need to be a vector of parameters, rather than having `mu1`, `mu2` etc like in the t-test models.

The Prior

- We have five parameters: a `mu` for each location, and a `sigma` that applies to all locations

- We will use a *hierarchical model*. There will be zero probability for all means being exactly equal, but there will be probability they are *similar*

Hierarchical Model Prior

```
# Priors for the four means
for(i in 1:4)
{
  mu[i] ~ dnorm(grand_mean, 1/diversity^2)
}

# and the hyperparameters
grand_mean ~ dnorm(0, 1/1000^2)
log_diversity ~ dunif(-10, 10)
diversity <-  exp(log_diversity)
```

Running the model

Let's look at:

- Trace plots for the `mu`s
- Posterior prob that `mu[2]` > `mu[3]`
- Posterior distribution of `log_diversity`

Pre-whitening

- Notice how the *priors* are correlated, and remember that JAGS is less efficient when the *posterior* is correlated.
- "Pre-whitening" can help: write your priors so they are independent.

```
# Old version of p(parameters | hyperparameters)
for(i in 1:4)
{
  mu[i] ~ dnorm(grand_mean, 1/diversity^2)
}


# Pre-whitened version
for(i in 1:4)
{
  n[i] ~ dnorm(0, 1)
  mu[i] <- grand_mean + diversity*n[i]
}
```

Part 2 – Model Checking

Ed Jaynes Quote

"It is as true in probability theory as in carpentry that introduction of more powerful tools brings with it the obligation to exercise a higher level of understanding and judgment in using them. If you gave a carpenter a fancy new power tool, he may use it to turn out more precise work in greater quantity; or he may just cut off his thumb with it. It depends on the carpenter."

The good news

- Bayesian Statistics always works! :-D

- ***IF*** your prior and likelihood are a good description of the prior beliefs and the experiment, and you can do the calculations correctly!!!

The bad news

For silly choices of prior and/or likelihood, it is quite possible to get a silly posterior distribution.

In practice, choices are often based on convenience and tradition.

Spherical Cow in a Vacuum

- http://en.wikipedia.org/wiki/Spherical_cow

Spherical Cow in a Vacuum

- http://en.wikipedia.org/wiki/Spherical_cow

- Commonly, people build their Bayesian models with "out of the box" standard probability distributions. Not *always* sensible.

- Outliers are one example that we have studied (and we saw one way of handling them).

Let's Be Concrete

- We will use a *linear regression* example

Making Some Fake Data

$$
\begin{aligned}
y_i &= \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \epsilon_i \\
\epsilon_i &\sim \mathcal{N}(0, \sigma^2)
\end{aligned}
$$

Simulation has a quadratic term
But we will fit the data without it...

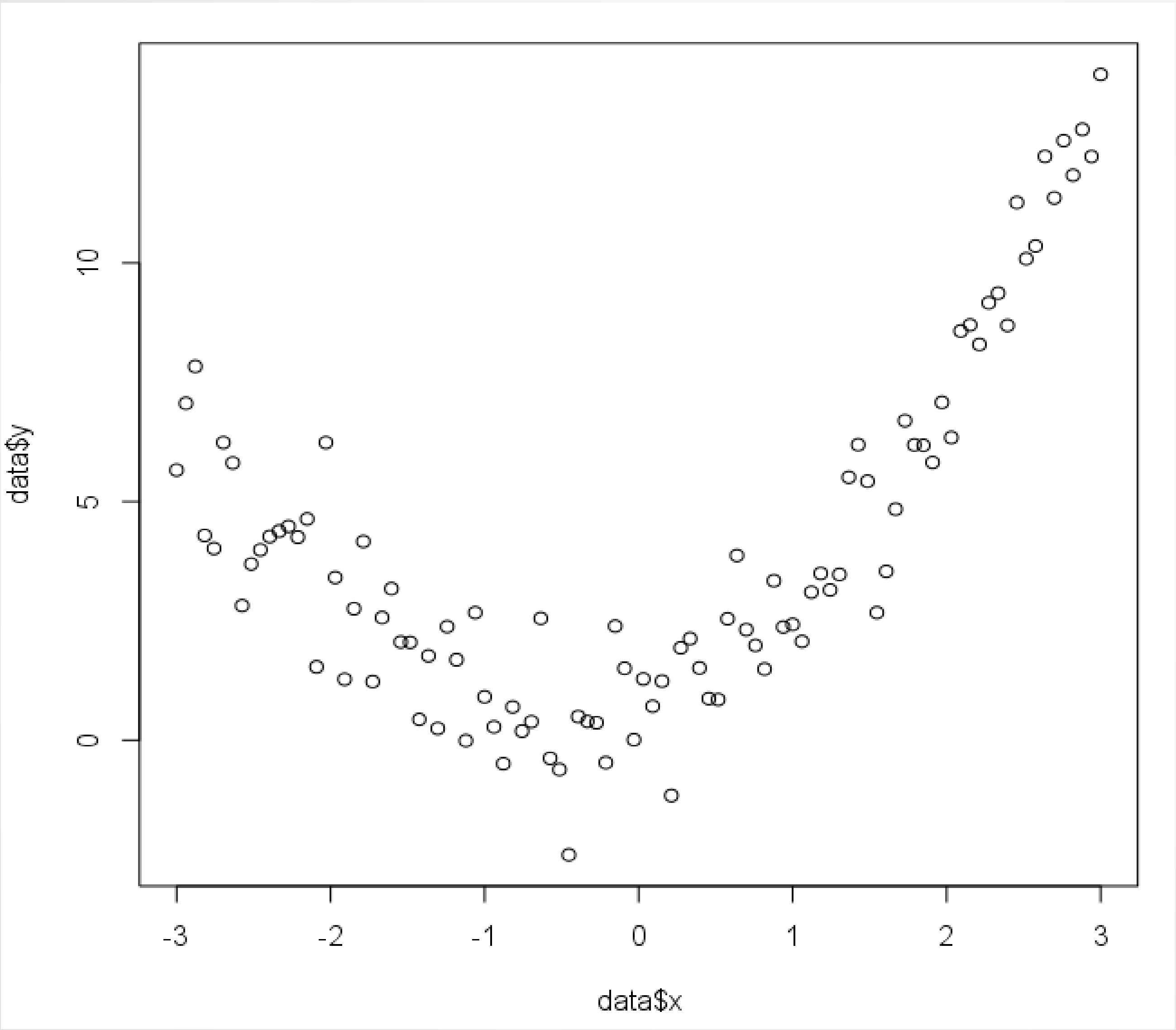
Making Fake Data in R

```
x = seq(-3, 3, length=100)
y = 1 + x + x^2  # Quadratic dependence
y = y + rnorm(100)   # Add noise


data = list(x=x, y=y, N=100)


# By the way...
dump('data', 'filename.r') # Save list to a
  'sourceable' file
```

Today's Fake Data



Generated from a Quadratic relationship

Common Sense

- Look at the data

- Can tell a linear fit is inappropriate

Question: Does "looking at the data" before deciding on your model contradict Bayesian principles?

Simple Linear Regression

```
model
{
    beta0 ~ dnorm(0, 1/1000^2)
    beta1 ~ dnorm(0, 1/1000^2)
    log_sigma ~ dunif(-10, 10)
    sigma <- exp(log_sigma)
    for(i in 1:N)
    {
        y[i] ~ dnorm(beta0 + beta1*x[i],
  1/sigma^2)
    }
}
```

Residuals

- Remember when doing a classical linear regression, you can look at the **residuals**

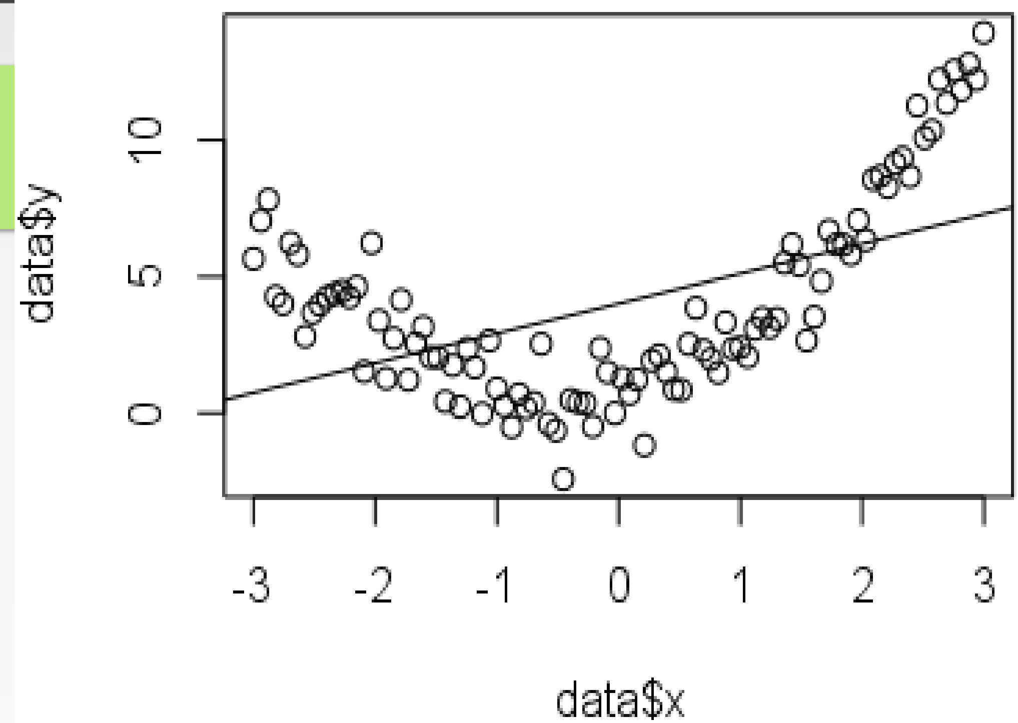
- **Can do the same here, but there's not just one set of residuals to look at**

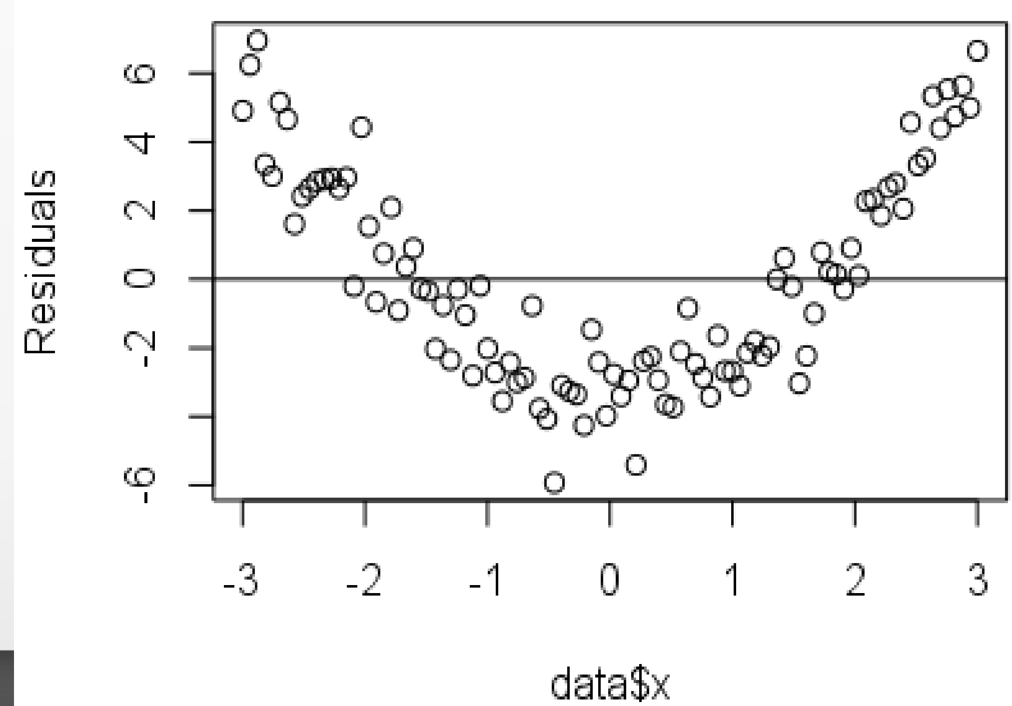
Residuals

```
lm(data$y ~ data$x)
```

Usual language: check residuals for “normality”

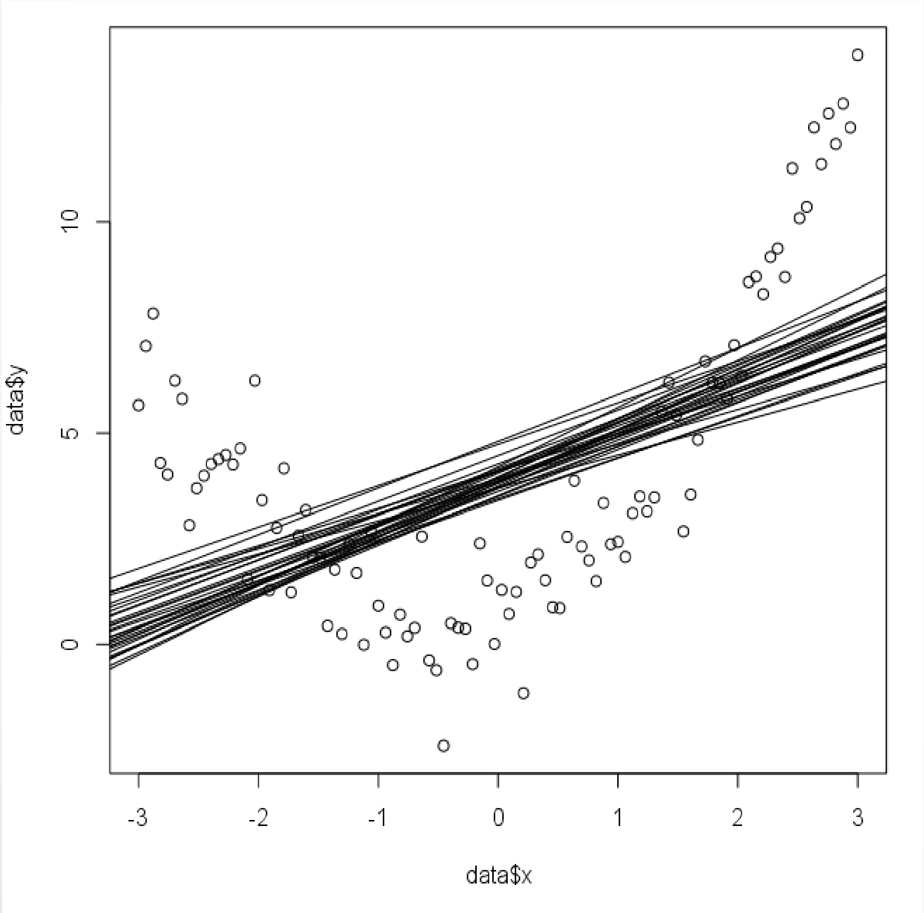
Rant opportunity, if time available :-)





Residuals – Bayesian

A posterior distribution for straight lines
Therefore, a posterior distribution for the residuals



Movie

- YouTube time

http://www.youtube.com/watch?v=ez0d5wlZ4rc

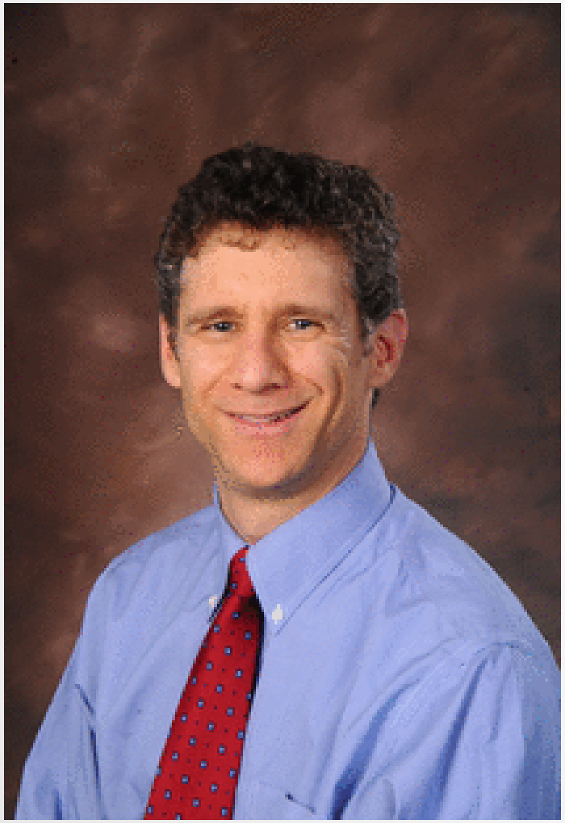
Residuals

- Looking at residuals is great, if it makes sense.

- It usually makes sense when the likelihood is something like

```
y[i] ~ dnorm(mu[i], 1/sigma^2)
```

Posterior Predictive Checks



Andrew Gelman
(Columbia University, NY)

Posterior predictive checks can be done for any kind of model, even ones where there isn't an obvious notion of “residuals”

Posterior Predictive Checks

- This is a somewhat sensible technique for model checking, and a good sanity check
- Take posterior distribution for parameters and use it to predict new data
- Simulated data should “resemble” actual data!

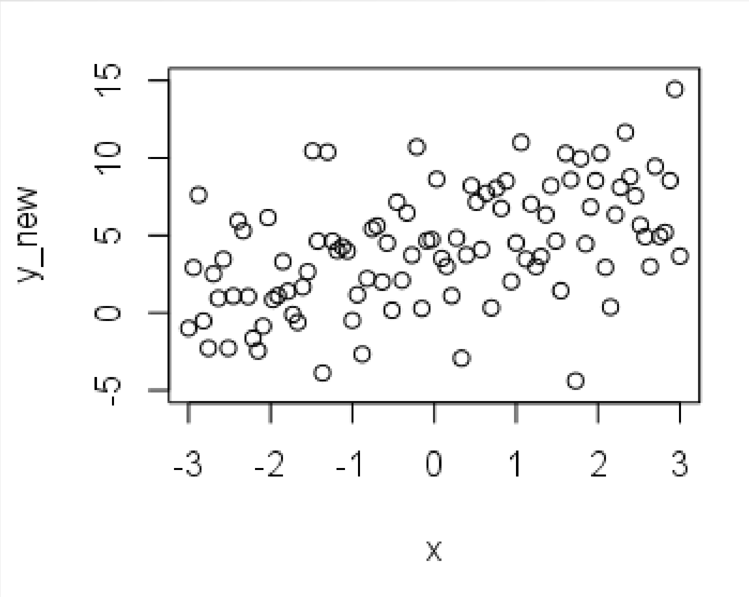
Simulating New Data

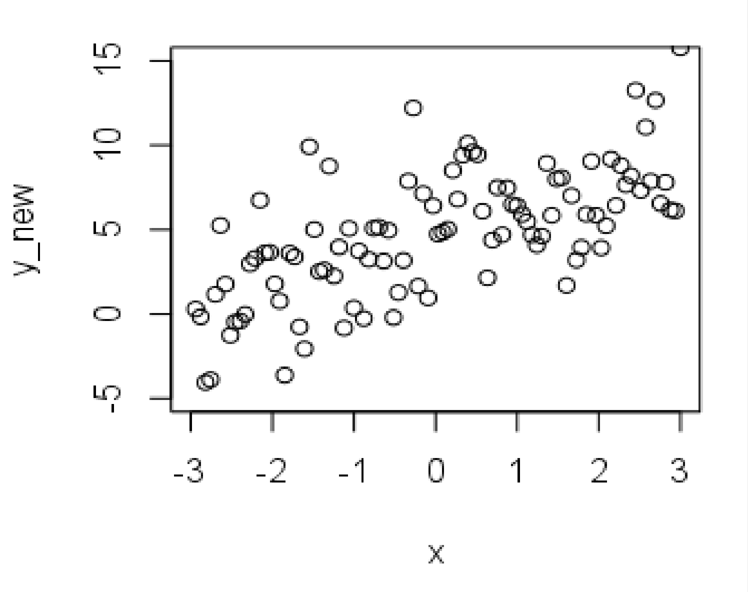
```
# Likelihood
for(i in 1:N)
{
  mu[i] < - beta0 + beta1*x[i]
  y[i] ~ dnorm(mu[i], 1/sigma^2)
  y_new[i] ~ dnorm(mu[i],
   1/sigma^2)
}
```

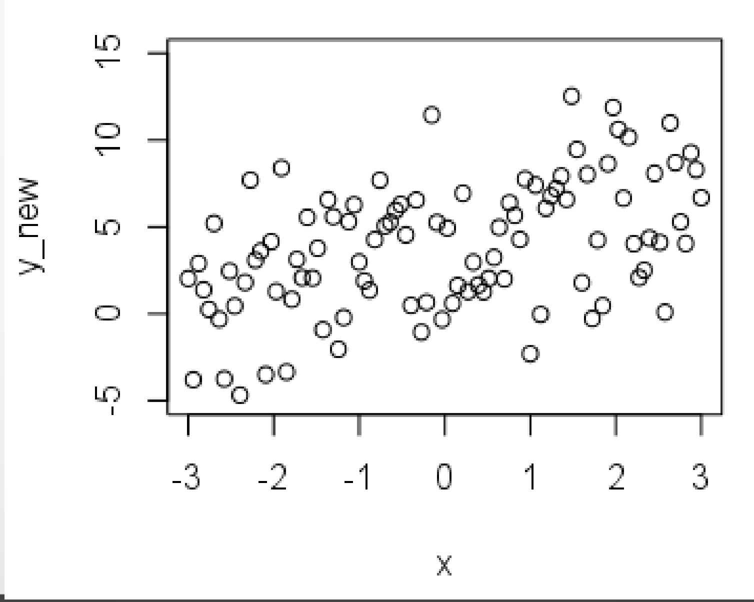
- In JAGS, simulated replicate data is basically another copy of the likelihood!

- But it's unobserved (whereas the actual data is observed, since it can be found in the `data` list)

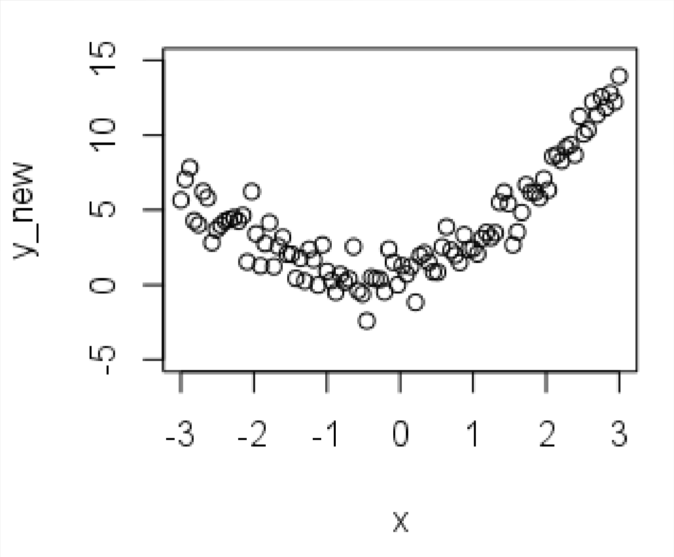
Simulated Replicate Data Sets

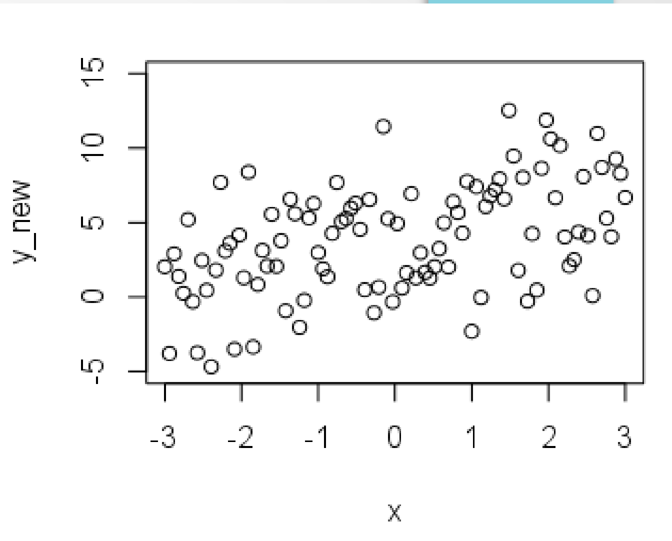


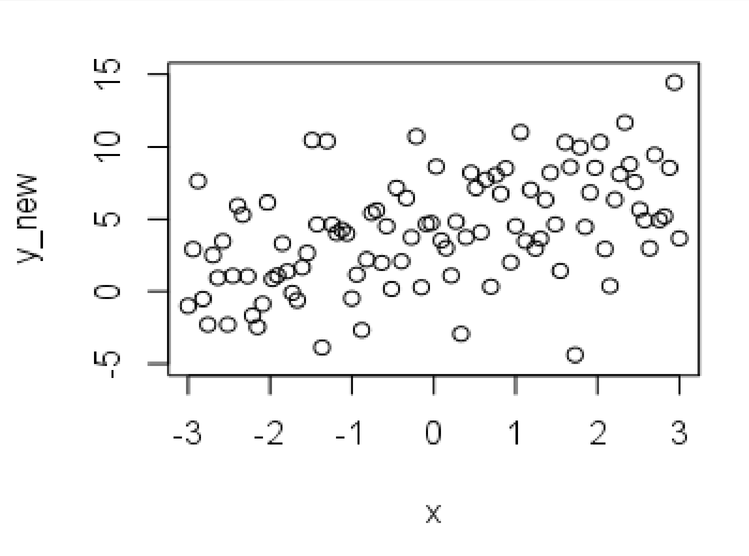


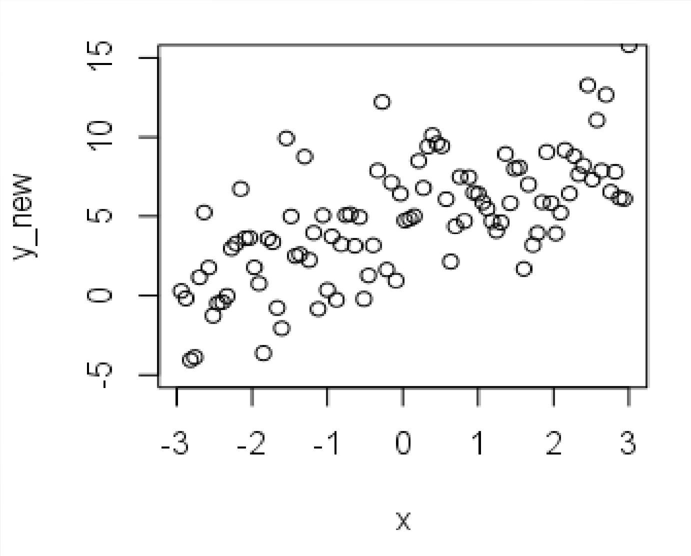


Which One is the Actual Data?









Key Points 1

- The actual data should probably not be an *outlier* from the distribution of data you would simulate from the model

Passing a PPC: No guarantee your inferences are good!

Failing a PPC: Doesn't prove that your inferences are bad!

So why do it?

Key Points 2

So why do it?

– So you understand your model better, and aren't accidentally incorporating unrealistic assumptions without realising it

Example from my research