

Reproducible Research Demo: What Does This Code Do Again?

Ben Chan^{1*}, Stephanie Renfro¹, Thomas Meath^{1,2}

1 Center for Health Systems Effectiveness, Oregon Health & Science University, Portland, Oregon, USA

2 Center to Improve Veteran Involvement in Care, Portland VA Medical System, Portland, OR, USA

*** E-mail: chanb@ohsu.edu**

Abstract

Reproducible research is the idea that results are published along with documentation of the coding steps that produced them. This additional transparency fosters credibility and allows others to replicate the findings and build upon them - and allows you to return to past work with less head scratching! The need for reproducibility is only increasing as datasets grow larger and data analyses become more complex.

Author summary

The authors are all Research Associates at OHSU's Center for Health Systems Effectiveness

Introduction

This work was inspired by the following email from Farmer Ben.

From: Ben Chan
Sent: Thursday, June 11, 2015 4:04 PM
To: Stephanie Renfro
Subject: What to feed chicks

Hello,

I'm receiving 20 baby chicks next month. Can you help me decide what to feed them? I'm choosing between the following four diets:

1. Grower diet
2. Layer diet
3. Breeder diet
4. High cluckage diet

Thanks, Ben

Ben Chan, Farmer and Research Associate
OHSU Center for Health Systems Effectiveness
Office: 3030 SW Moody | Mail Code: MDYCHSE
www.ohsu.edu/chse

Results

Results and Discussion can be combined.

Preliminaries

Start clock to calculate total runtime.

```
start_program <- proc.time()
```

Load needed packages:

- *data.table* - for faster processing
- *knitr* - for better tables (“kable” function)
- *ggplot2* - for pretty plots
- *knitr* - for better table display

```
packages <- c("data.table", "ggplot2", "knitr")
sapply(packages, require, character.only=TRUE, quietly=TRUE)
```

```
data.table    ggplot2    knitr
      TRUE      TRUE      TRUE
```

Define the CHSE color palette function.

```
colorPalette <- function () {
  c(rgb( 1,  67, 134, maxColorValue=255),
    rgb(119, 120, 123, maxColorValue=255),
    rgb(139, 184, 234, maxColorValue=255),
    rgb(188, 190, 192, maxColorValue=255),
    rgb( 94, 122, 162, maxColorValue=255),
    rgb(223, 122,  28, maxColorValue=255))
}
```

Prepare Data

This demo uses data from an experiment on the effect of diet on early growth of chicks, *ChickWeight*, which comes pre-loaded in any R session.

Let’s take a look at the first few rows:

```
head(ChickWeight)
```

```
  weight Time Chick Diet
1     42    0     1    1
2     51    2     1    1
3     59    4     1    1
4     64    6     1    1
5     76    8     1    1
6     93   10     1    1
```

Let's also print a summary of the data.

Note, by specifying the option "echo = FALSE", the resulting output will display, but not the code that generated it.

weight	Time	Chick	Diet
Min. : 35.0	Min. : 0.00	13	: 12 1:220
1st Qu.: 63.0	1st Qu.: 4.00	9	: 12 2:120
Median :103.0	Median :10.00	20	: 12 3:120
Mean :121.8	Mean :10.72	10	: 12 4:118
3rd Qu.:163.8	3rd Qu.:16.00	17	: 12
Max. :373.0	Max. :21.00	19	: 12
		(Other):506	

Convert to data.table for faster processing.

```
ChickWeight <- data.table(ChickWeight)
```

Just for fun, let's create a table showing mean weight at times 0, 10, and 21 days, for each of the four diet types.

```
mean_ChickWeight <- ChickWeight[Time %in% c(0,10,21),
                                list(mean_weight = round(mean(weight), digits=1)),
                                by = list(Diet,Time)]
```

```
#kable(mean_ChickWeight)
```

Create a character variable for diet. Use this variable for plotting small multiples.

```
ChickWeight[, dietChr := sprintf("Diet %d", Diet)]
```

	weight	Time	Chick	Diet	dietChr
1:	42	0	1	1	Diet 1
2:	51	2	1	1	Diet 1
3:	59	4	1	1	Diet 1
4:	64	6	1	1	Diet 1
5:	76	8	1	1	Diet 1

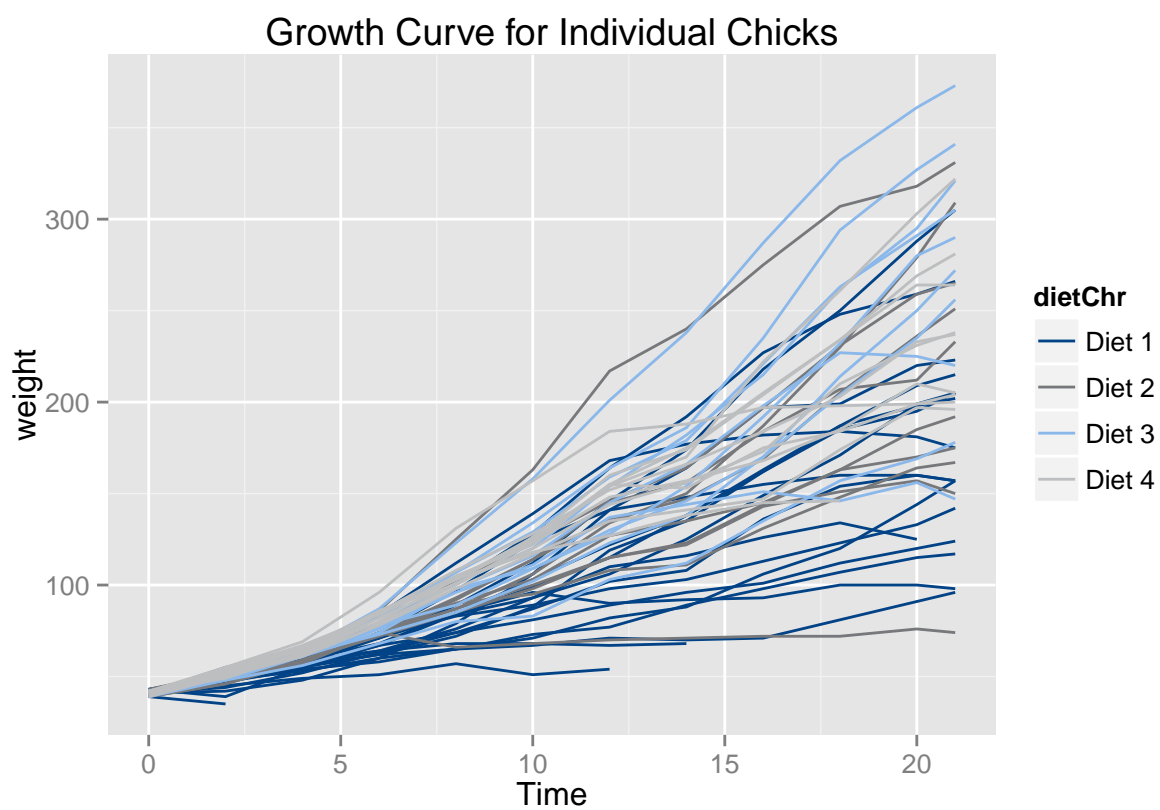
574:	175	14	50	4	Diet 4
575:	205	16	50	4	Diet 4
576:	234	18	50	4	Diet 4
577:	264	20	50	4	Diet 4
578:	264	21	50	4	Diet 4

Growth for Individual Chicks

The following plot illustrates the growth curve for individual chicks from 0 to 21 days.

Colors represent the four diets.

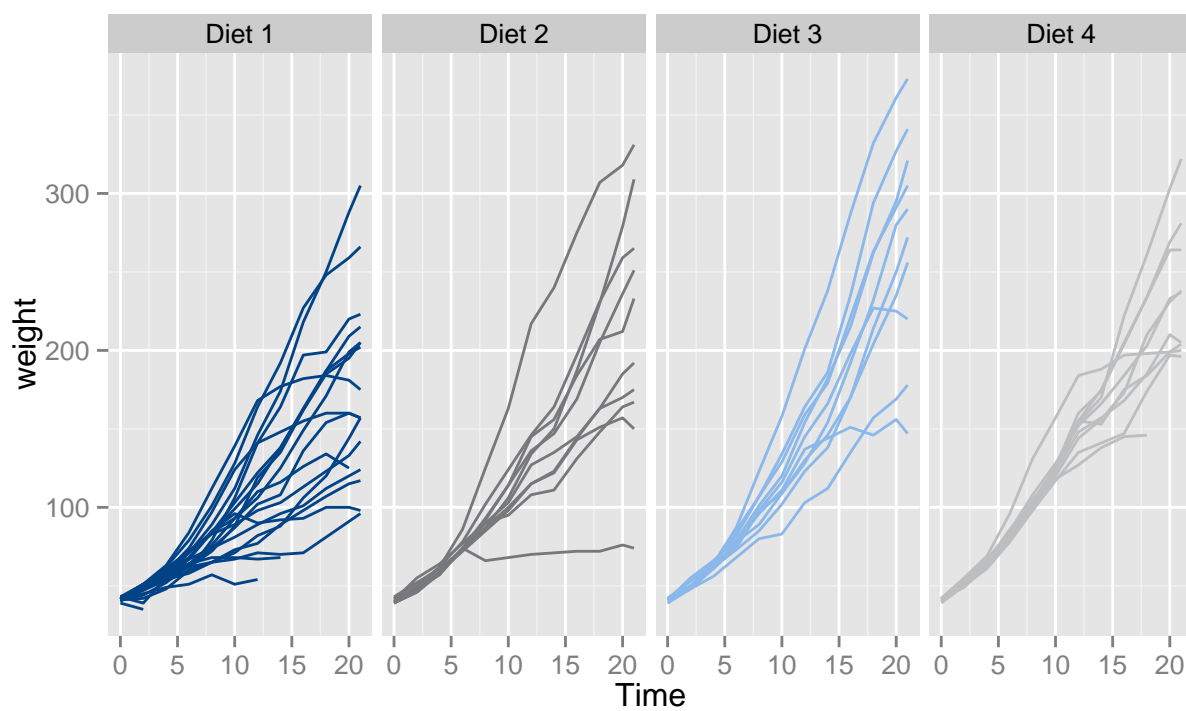
From this plot, it is difficult to distinguish between the performance of the four diets.



Individual growth curves

Plot individual chick growth curves using small multiples.

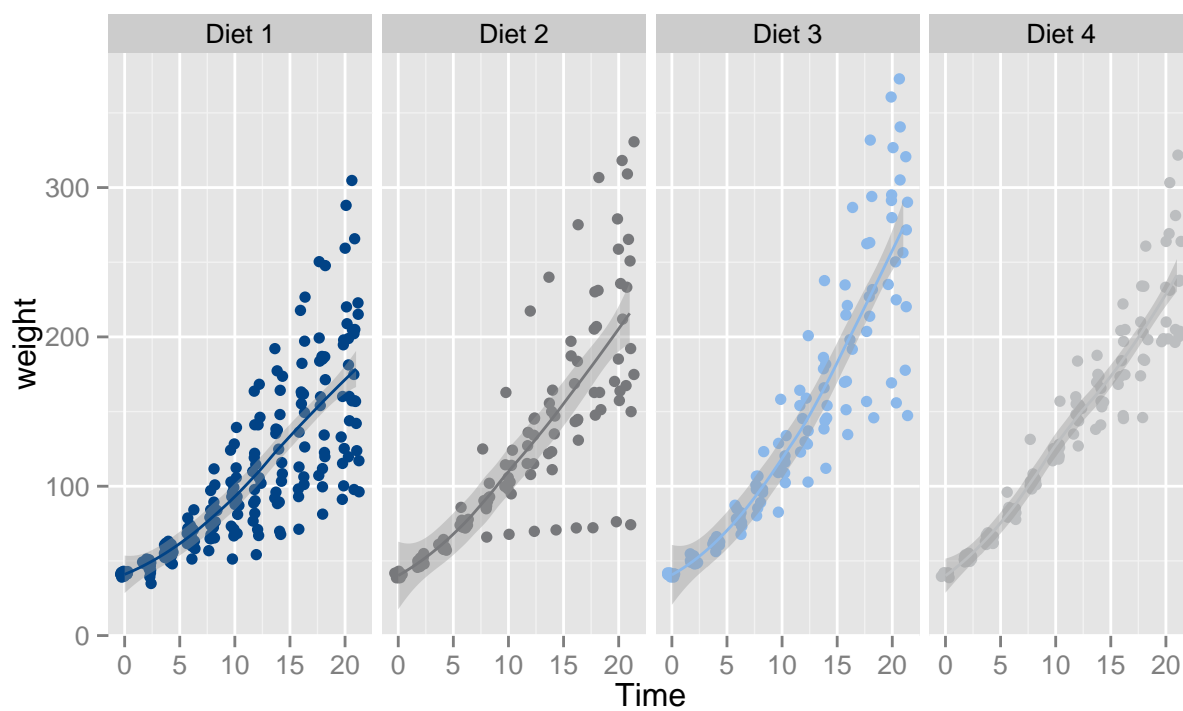
Growth Curve for Individual Chicks



Fitted growth curves

Plot fitted growth curves using small multiples. Data points are jittered around time value.

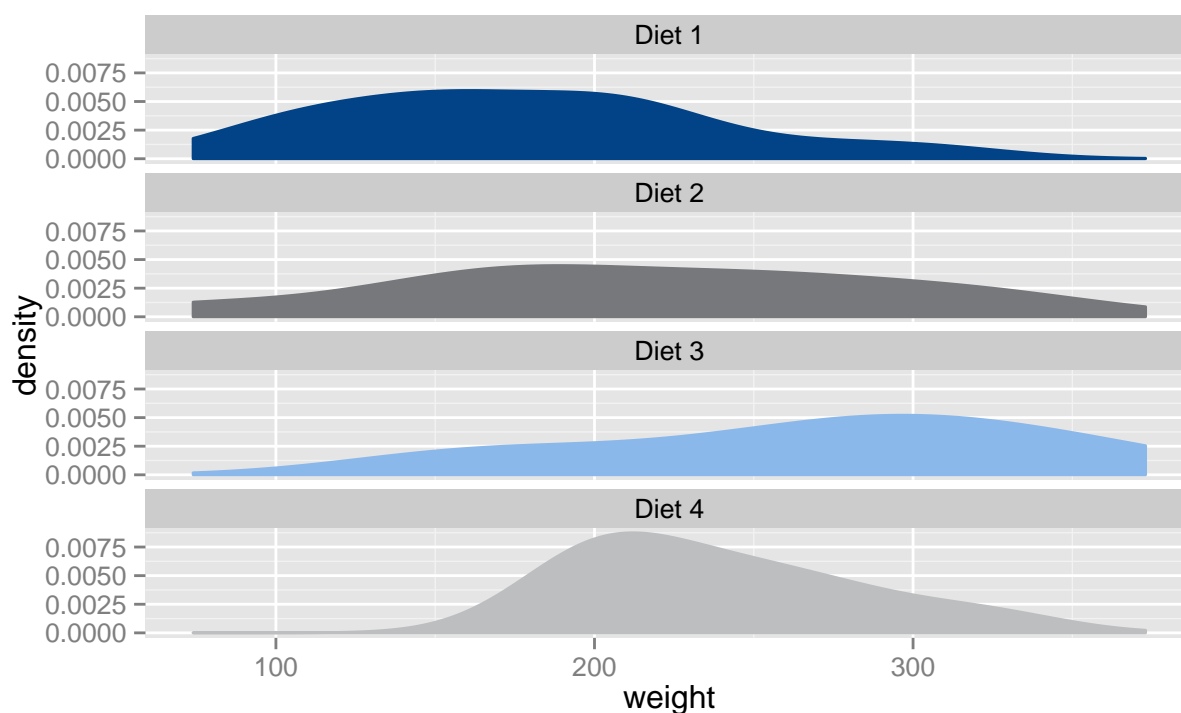
Fitted Growth Curves



Final weight density

Plot densities by diet for chicks' final weights (day 21) using small multiples.

Density: Final Weight



Wrap Up

Calculate total runtime.

```
time_program <- proc.time()-start_program
print(paste("Total runtime:", format(time_program[3]/60,digits=3), "minutes"))
```

```
[1] "Total runtime: 0.0852 minutes"
```

Clear memory.

```
rm(list=ls())
invisible(gc())
```

Discussion

If this was a real paper, we would have some discussion about our results. Because this is an example, we do not.

Methods

This section is also not currently written out in full. This is where your methods would go to satisfy PLOS.

Acknowledgments

Rojo the Llama: for being a very friendly llama.

References

A reference list should be automatically created here. However it won't. Pandoc will place the list of references at the end of the document instead. There are no convenient solution for now to force Pandoc to do otherwise. The easiest way to get around this problem is to edit the LaTeX file created by Pandoc before compiling it again using the traditional LaTeX commands.

Figure Legends

Figure Legends would go in here

Tables

Tables would go here: