# Honey, I Shrunk the Bees!

Chris G Martin

May 15, 2016

# Running of the Bees

- Bees: we love 'em, yet we don't help 'em
  - The most **effective** polinators on the planet
  - Our history dates back over 8,000 years ago
- Honey helps our economy and our health
  - Used in desserts, breads, barbecues, mustards, jellies and ointment treatments
  - Loaded with antibacterial and antifungal properties
  - Can treat dandruff, is used in energy drinks, and can treat wounds and burns
  - Can also help fight local allergies

# Running of the Bees

- Bees: we love 'em, yet we don't help 'em
  - The most **effective** polinators on the planet
  - Our history dates back over 8,000 years ago
- Honey helps our economy and our health
  - Used in desserts, breads, barbecues, mustards, jellies and ointment treatments
  - Loaded with antibacterial and antifungal properties
  - Can treat dandruff, is used in energy drinks, and can treat wounds and burns
  - Can also help fight local allergies
- **But do we really on them?**

# Overview

- Data
  - Sources, Manipulation, Cleaning, and Tidying
  - Exploration
- Inference
  - Summarising the Data
  - Predicting Honey Production
- Analysis
  - Linear Regression
  - Multiple Linear Regression
- Wrap Up
  - Conclusion
  - Next Steps

# Data: Sources, Manipulation, Cleaning, and Tidying

- NASS: National Agriculture Statistics Service

# Data: Sources, Manipulation, Cleaning, and Tidying

- NASS: National Agriculture Statistics Service
- MySQL: Data Imported and Manipulated in MySQL

| Program | Year | Period | State | DataItem | Value |
|---|---|---|---|---|---|
| **CENSUS** | 2014 | YEAR | AL | HORTICULTURE TOTALS - OPERATIONS WITH SALES | 3.332205 |
| **CENSUS** | 2014 | YEAR | AL | HORTICULTURE TOTALS - OPERATIONS WITH SALES | 3.496508 |
| **CENSUS** | 2014 | YEAR | AL | HORTICULTURE TOTALS - OPERATIONS WITH SALES | 3.784190 |

# Data: Sources, Manipulation, Cleaning, and Tidying

- NASS: National Agriculture Statistics Service
- MySQL: Data Imported and Manipulated in MySQL
- Cleaning:
  - Most data was cleaned using MySQL
  - Data was then imported into R for further cleaning

```
#converted characters to values
honey_county2$Value <- as.integer(honey_county2$Value)
#changed NAs to 0
honey_county2$Value[is.na(honey_county2$Value)] <- as.integer(0)
#converted CountyANSI codes to 3 digits
honey_county2$CountyANSI <- sprintf("%03d", honey_county2$CountyANSI)
#converted 3 digit CountyANSI codes to 5 digits with StateANSI codes
honey_county2$CountyANSI <- as.numeric(paste0(honey_county2$StateANSI,
                                              honey_county2$CountyANSI))
```

# Data: Sources, Manipulation, Cleaning, and Tidying

- NASS: National Agriculture Statistics Service
- MySQL: Data Imported and Manipulated in MySQL
- Cleaning: Necessary conversions and ANSI code merging
- Tidying:
    - Tables were sorted by State name
    - Tables were transposed for analysis

```
#alphabatizing based on DataItem and State
honey_county2 %>%
  arrange(., DataItem) %>%
  arrange(., State)

#Transposing table by DataItem
honey_county2[, c(2,4,5,6,7,9,10)] %>%
  spread(., DataItem, Value)
```

# Data: Exploration

- Exploring Honey Production per Year

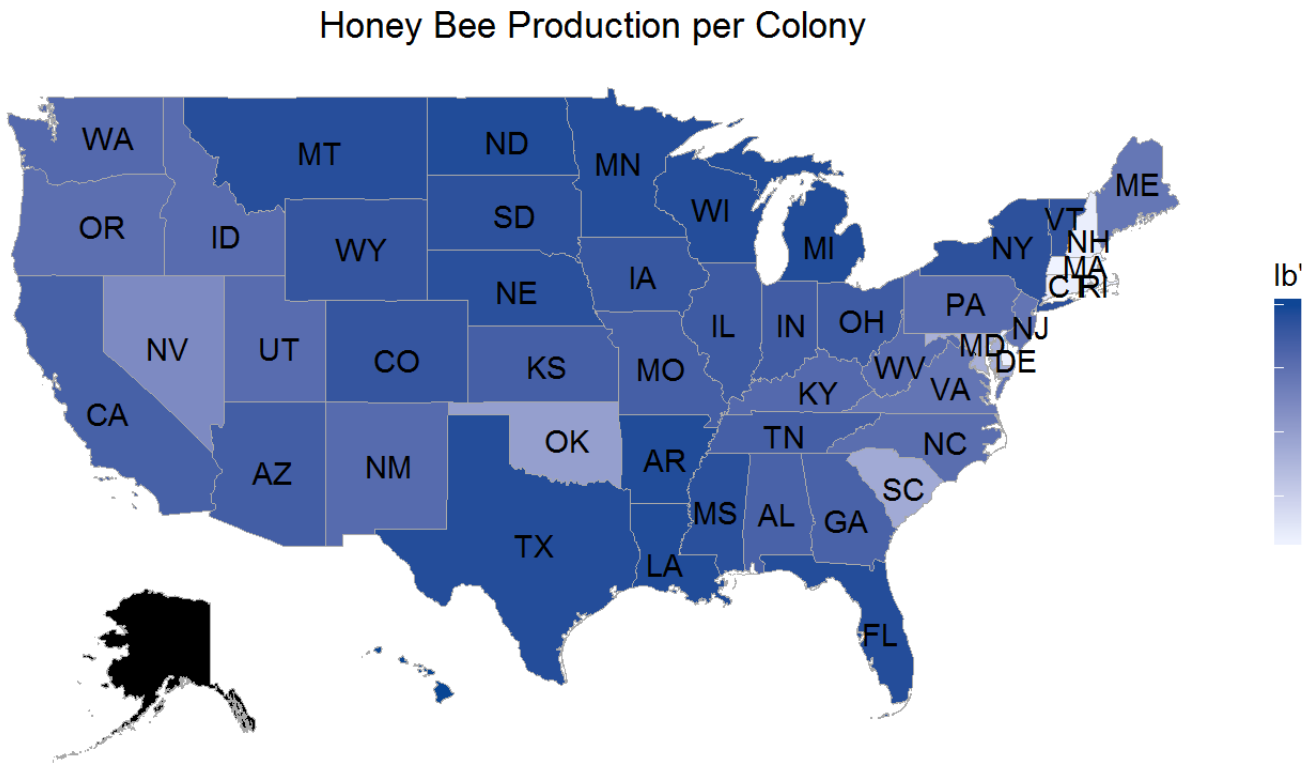
```
kable(tail(honey_state2[which(honey_state2$DataItem ==
                                'HONEY - PRODUCTION, MEASURED IN LB / COLONY'
                        c(3,10,13)] %>%
  group_by(Year) %>%
  summarise(value = sum(exp(Value)))))
```

| Year | value |
|---|---|
| 2010 | 1582.785 |
| 2011 | 1444.734 |
| 2012 | 1479.247 |
| 2013 | 1286.598 |
| 2014 | 1513.759 |
| 2015 | 1341.196 |

# Data: Exploration

- Exploring Bee Colonies
- ChoroplethR Maps:

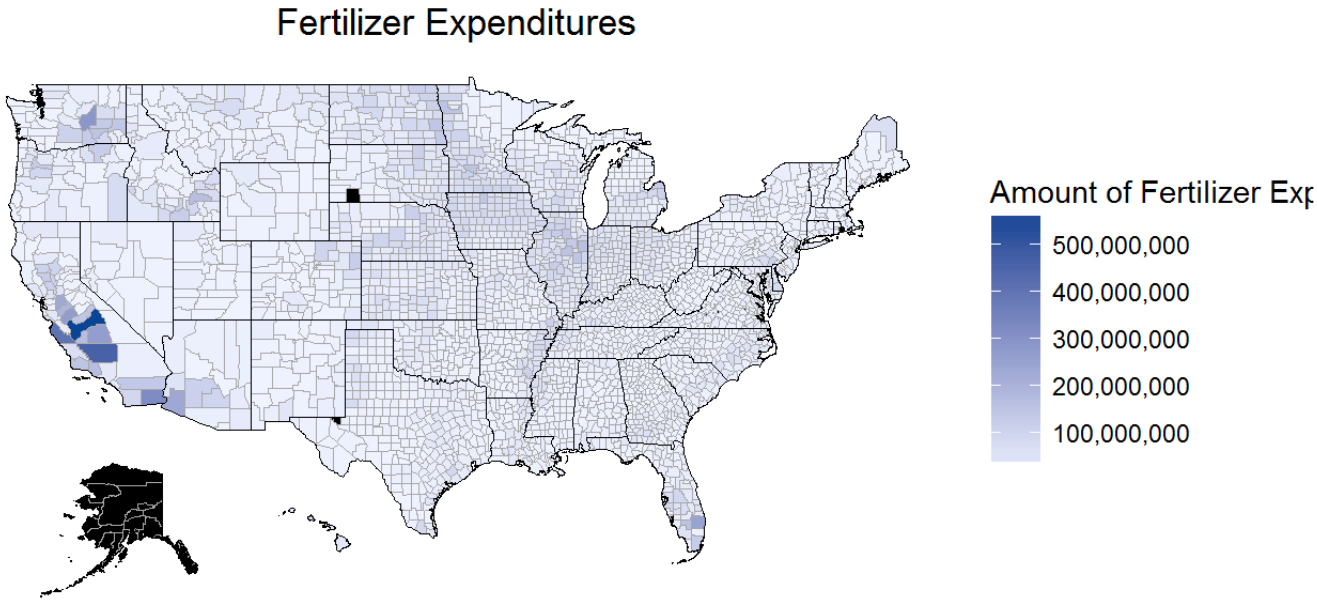
```
state_choropleth(HS_HonProdPlot, title = "Honey Bee Production per Colony",
                 legend = "lb's of Honey", num_colors = 1)
```



# Data: Exploration

- Exploring Bee Colonies
- ChoroplethR Maps:

```
county_choropleth(HC_FertPlot, title = "Fertilizer Expenditures", legend = "A
```



# Data: Exploration

- Exploring Bee Colonies
- ChoroplethR Maps
- Exploration Summary:
  - The number of honey bees has been dynamic in that the number of colonies decreased and increased
  - Some states have more expenditures in chemicals and fertilizers than others
  - There are plenty of acres of open horticulture in production for our wonderful bees to pollinate.

# Inference

- Summarising the Data:
  - Summary statistics on production of honey per colony

```
summary(honey_state3$'HONEY...PRODUCTION..MEASURED.IN.LB...COLONY')
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   2.000   3.000   4.000   3.562   4.000   5.000
```

# Inference: Production of Honey per Colony Changed from 1987 to 2015

```
plot_ly(StateProd2, x = Year, y = value, text = paste("Year: ", Year),
        mode = "markers", color = value, size = value)
```

# Inference

- Summarising the Data
- Predicting Honey Production:
    - Confidence Intervals

```
c(lower_vector[1], upper_vector[1])
```

```
## [1] 3.331427 3.788573
```

```
mean(population2)
```

```
## [1] 3.469388
```

```
mean(population3)
```

```
## [1] 3.365854
```

# Inference

- Summarising the Data
- Predicting Honey Production:
  - Confidence Intervals
  - Hypothesis Testing

```
inference(y = population6$Value, x = population6$Year, est = "mean",
          type = "ht", null = 0, alternative = "twosided", method = "theoreti
```

```
## Response variable: numerical, Explanatory variable: categorical
## Difference between two means
## Summary statistics:
## n_1987 = 49, mean_1987 = 3.4694, sd_1987 = 0.581
## n_2015 = 41, mean_2015 = 3.3659, sd_2015 = 0.4877
```

```
## Observed difference between means (1987-2015) = 0.1035
##
## H0: mu_1987 - mu_2015 = 0
## HA: mu_1987 - mu_2015 != 0
## Standard error = 0.113
## Test statistic: Z =  0.919
## p-value =  0.358
```

# Analysis

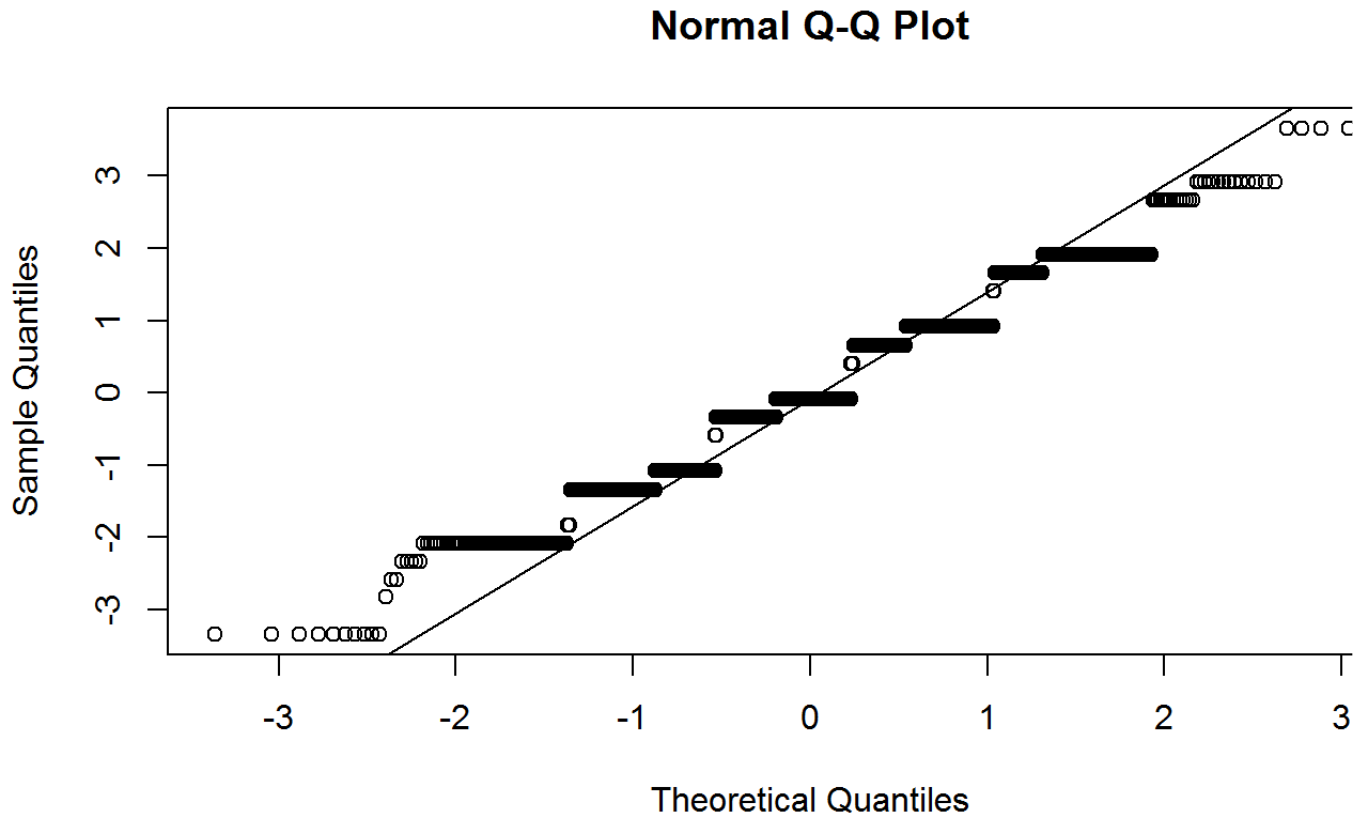
- Linear Regression:
  - Determining a casual relationship between the number of colonies and the production of honey per colony

$$\hat{lbsofhoneyproducedpercolony} = -7.10151 + 0.74615 * colonies$$

# Analysis

- Linear Regression

```
qqnorm(hc1$residuals)
qqline(hc1$residuals)
```




*#plot on next slide*

# Analysis

- Linear Regression
- Multiple Linear Regression:
  - Full Model
  - Backward Elimination
  - Forward Selection

# Analysis: Linear Regression

```
summary(m_final)
```

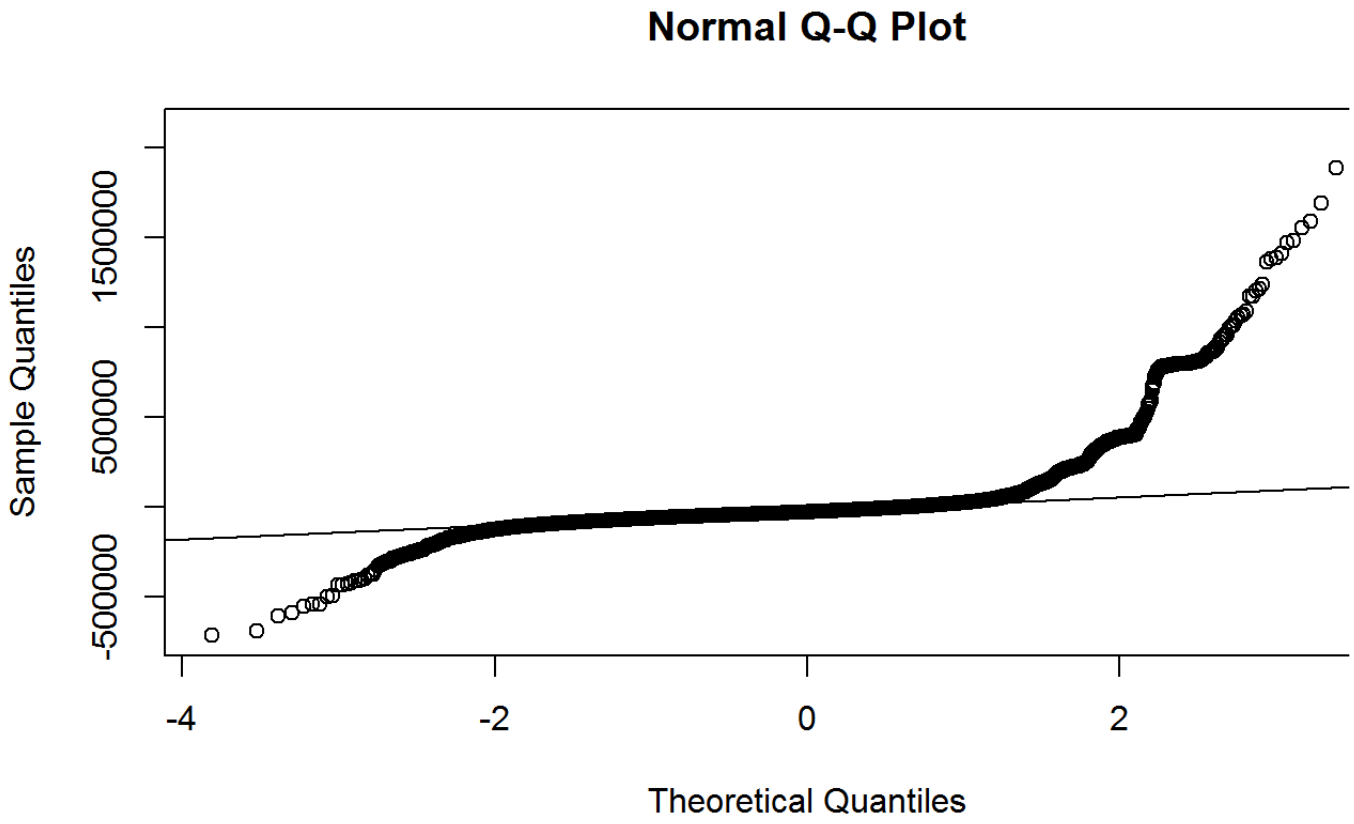
```
##
## Call:
## lm(formula = HoneyProd ~ HortUndProt + HoneySales + HoneyColSales +
##     ChemExp + HoneyOpsSales + CropOrgSales + HoneyOpsProd + HortExcTAcres
##     ChemOps + FertOps + FertExp + HortExcTIrgOps + CropSales +
##     HortExcTVSTOps + HortExcTIrgAcres + CropOps + HortExcTVSTSales,
##     data = HoneyC_FullModel)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -712524  -49322  -26912    2842 2104209
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)      -1.893e+04  2.068e+04  -0.915 0.360125
## HortUndProt       1.632e+03  3.614e+02   4.516 6.40e-06 ***
## HoneySales        4.913e-01  8.570e-03  57.334  < 2e-16 ***
## HoneyColSales     2.580e+02  1.761e+01  14.654  < 2e-16 ***
## ChemExp           2.940e-03  3.894e-04   7.550 4.92e-14 ***
## HoneyOpsSales    -7.200e+03  4.809e+02 -14.971  < 2e-16 ***
## CropOrgSales      1.271e-02  1.572e-03   8.087 7.12e-16 ***
## HoneyOpsProd      3.206e+03  3.597e+02   8.914  < 2e-16 ***
## HortExcTAcres    -4.177e+02  1.963e+02  -2.128 0.033337 *
## ChemOps           2.154e+02  2.385e+01   9.030  < 2e-16 ***
## FertOps          -2.032e+02  1.973e+01 -10.296  < 2e-16 ***
## FertExp          -1.821e-03  3.943e-04  -4.619 3.92e-06 ***
## HortExcTIrgOps    3.262e+01  8.001e+00   4.076 4.62e-05 ***
## CropSales         3.293e+03  1.364e+03   2.415 0.015766 *
## HortExcTVSTOps   -2.420e+02  7.189e+01  -3.367 0.000764 ***
## HortExcTIrgAcres -1.723e+01  6.100e+00  -2.824 0.004759 **
## CropOps           4.217e+01  1.695e+01   2.488 0.012886 *
## HortExcTVSTSales  2.195e-04  1.109e-04   1.978 0.047959 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

# Analysis

- Linear Regression
- Multiple Linear Regression:
  - Full Model
  - Backward Elimination
  - Forward Selection
  - Checking Model Assumptions

```
qqnorm(m_final$residuals)
qqline(m_final$residuals)
```

# Analysis: Multiple Linear Regression



# Conclusion: Plight of the Bumblebee

I fear that we are no closer to answering the question posed at the start: Do we rely on the bees? It is perhaps due to my selection of explantory variables, or heavy skewing of the data by using averages and 0's for missing values, or it could be simply that we were dealing with a time series (square peg in a round hole). Regardless, the analysis was very successful in uncovering some interesting information in the decline of honey production per colony since 1987. Our fears may be different than what I had expected. Rather than focusing our attention on the number of bees (which we can inconclusively say we should continue to worry about), **we should focus our attention on a possible exponential decline in honey.**

# Next Steps

There are several directions we could go from here. One direction could be analysing the number of bees in a colony and see how they influence the amount of honey a colony can produce: is there a point of marginal returns whereby adding bees to a colony actually hinders the production of honey? How good are the bees at pollinating and how efficient are they? In the end, my goal is to save the bees and help man-kind learn to live alongside them (and nature in general).

Another analysis would be to see the reverse of what I've done here: rather than see how these variables explain the honey production of a colony, see how honey production influences the other variables. This would give us much more insight into the over-arching question I posed at the outset: Do we rely on the bees?

# Thank You

- Contact
  - Chris G Martin
  - chrisgmartin2@gmail.com
- Resources
  - GitHub: https://github.com/chrisgmartin/HoneyIShrur
  - Full Final Project: http://rpubs.com/chrisgmartin/HISB
- Data Sources:
  - NASS: https://quickstats.nass.usda.gov
  - State Table: http://www.statetable.com