

# Reconstructing family size from juvenile percentage

From previous data exploration, we know that around 1200 records have family size data, ie, the numbers of families of each size. This leaves the majority of the dataset of some 5800 records without any such information. However, most of the data have a juvenile percentage, or if not, the absolute number of juveniles. Here, I'll try to reconstruct one field – total number of families – using the 1200 records available, in an effort to allow me to use the remainder of the records.

## **Family size counts: How does the total number of families of each size vary with zone and year?**

The numbers of families of each size is a power law decay. 50% more families with a single juvenile are seen than families with two, and 200% more than families of three, and around 500% more than families of 5. Thereon, larger families are negligibly rare. There appears to be some relation to the lemming cycle: in lemming-peak years, such as 2002 and 2008, there are fewer families with two or more juveniles. In lemming-peak years, such as 2005, and possibly 2001, the number of families with two juveniles is higher than those with one.

## **Family sums: How many families in total are present in flocks in the different zones and in different years?**

There doesn't appear to be as great an influence of the lemming cycle on the number of families. Possibly, the lemming-crash reduces the reproductive success of geese and results in smaller families, and also in fewer families.

---

## **Number of families $\sim$ number of juveniles**

Goose flocks are formed by the aggregation of families (W. H. Elder and Elder 1949), and this would mean that the number and size of families determines the number and percentage of juveniles in a flock. It should be possible to derive a relationship between the number of juveniles and the number of families in a flock.

The raw family sum data shows an increasing spread with  $x$ , and is then log transformed. This data now shows a saturating relationship with the number of juveniles. Log-log transforming the data shows a linear relationship without increasing spread with  $x$ .

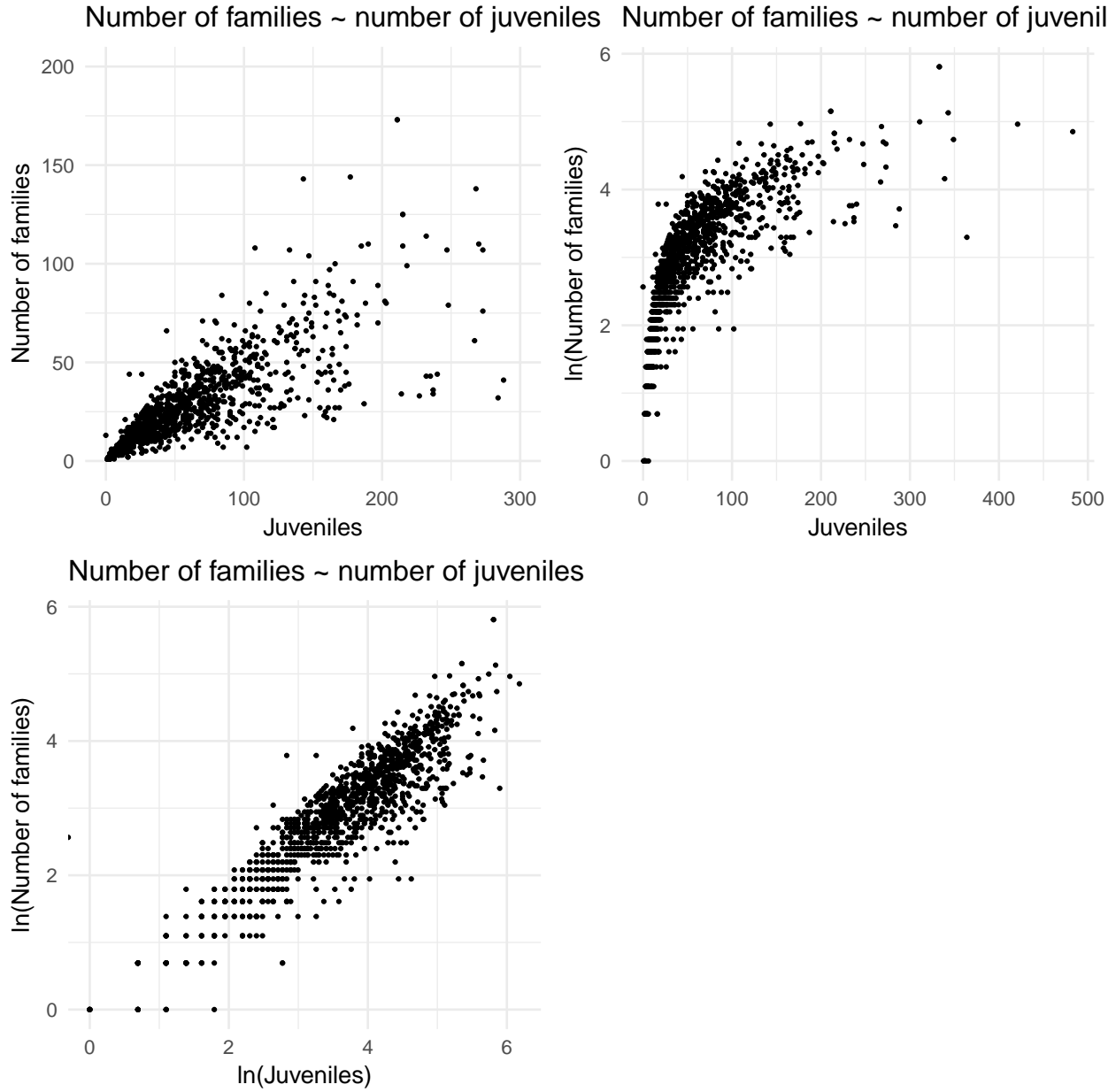


Figure 1: Number of families in a flock ~ number of juveniles in the flock. The log transformed sum of families shows a saturating relationship against the number of juveniles. The spread of the data is better constrained by the log-log transformation.

## Predicting total families from juvenile counts, non-linear model

I used the data with log-transformed total families and raw juvenile count to run a simple non-linear regression. Since the log total families shows saturation with increasing juvenile count, I chose to model the relationship between  $y = \log(\text{total families})$  and  $x = \text{juvenile count}$  with a simple Michaelis-Menten function,  $y = y_m \frac{x}{t_{1/2} + x}$  using the `nls` function built into R. Starting parameter estimates were set at  $y_m = 6$  and  $t_{1/2} = 50$ . The model appears to fit the data fairly well, with a correlation of 0.9046893 between the real and predicted values.

```
## Warning: Removed 44 rows containing non-finite values (stat_smooth).
```

```
## Warning: Removed 44 rows containing missing values (geom_point).
```

## Predicting total families from juvenile counts, linear model

```
## Warning: Removed 45 rows containing non-finite values (stat_smooth).
```

```
## Warning: Removed 44 rows containing missing values (geom_point).
```

I also tried an alternative approach, using log-log transformed data on total families and juvenile counts to run a simple linear model. The model fits the data and appears to explain much of the variation (adjusted R-squared = 0.8319876).

## Comparing models

A preliminary comparison of the models appears to show no significant difference. The model using log-log transformed data has a lower AIC score (1161.1260676) than the log-total family model (1257.8109035).

Number of families ~ number of juveniles, log transform

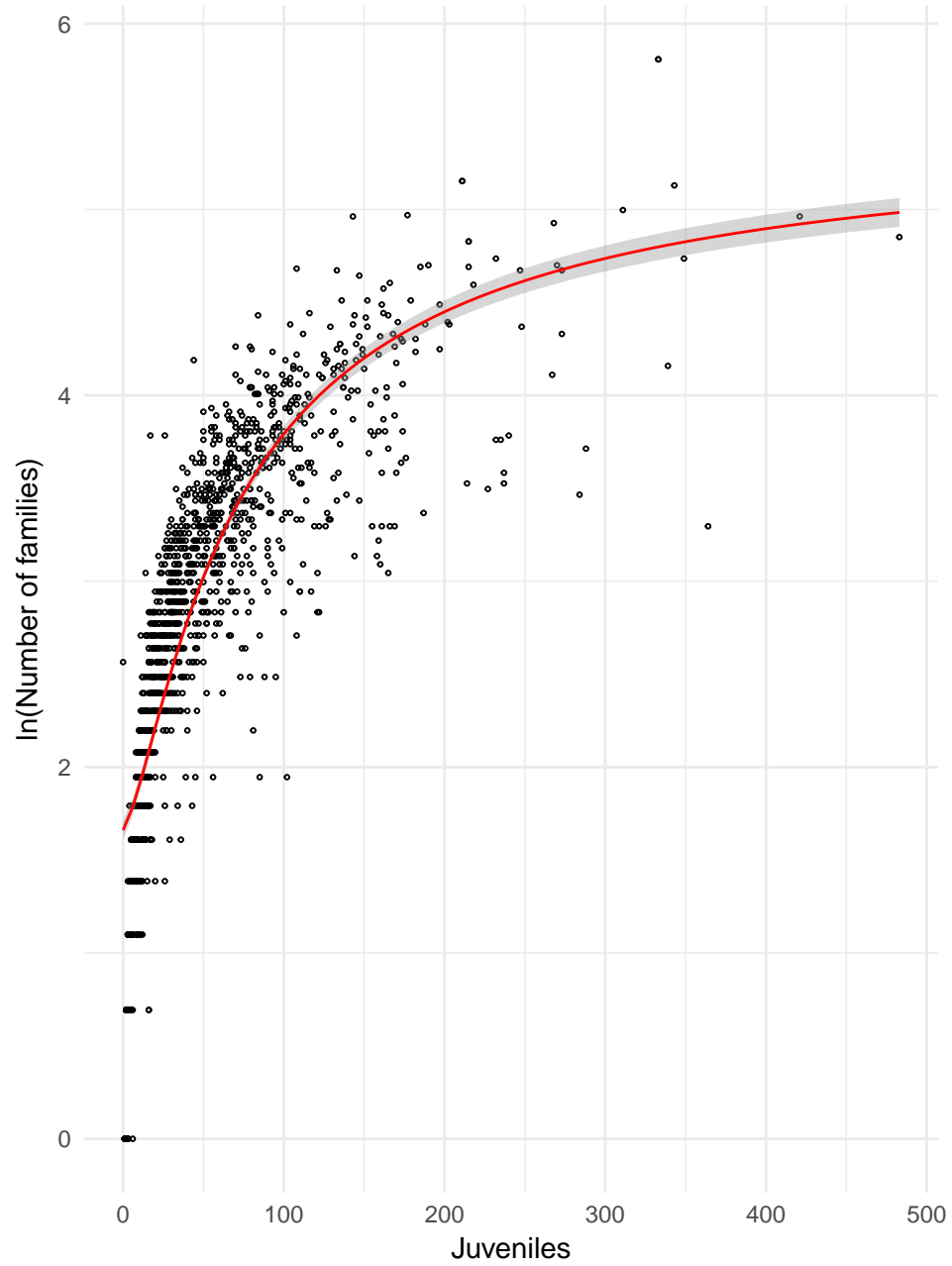


Figure 2: Model fit and data.

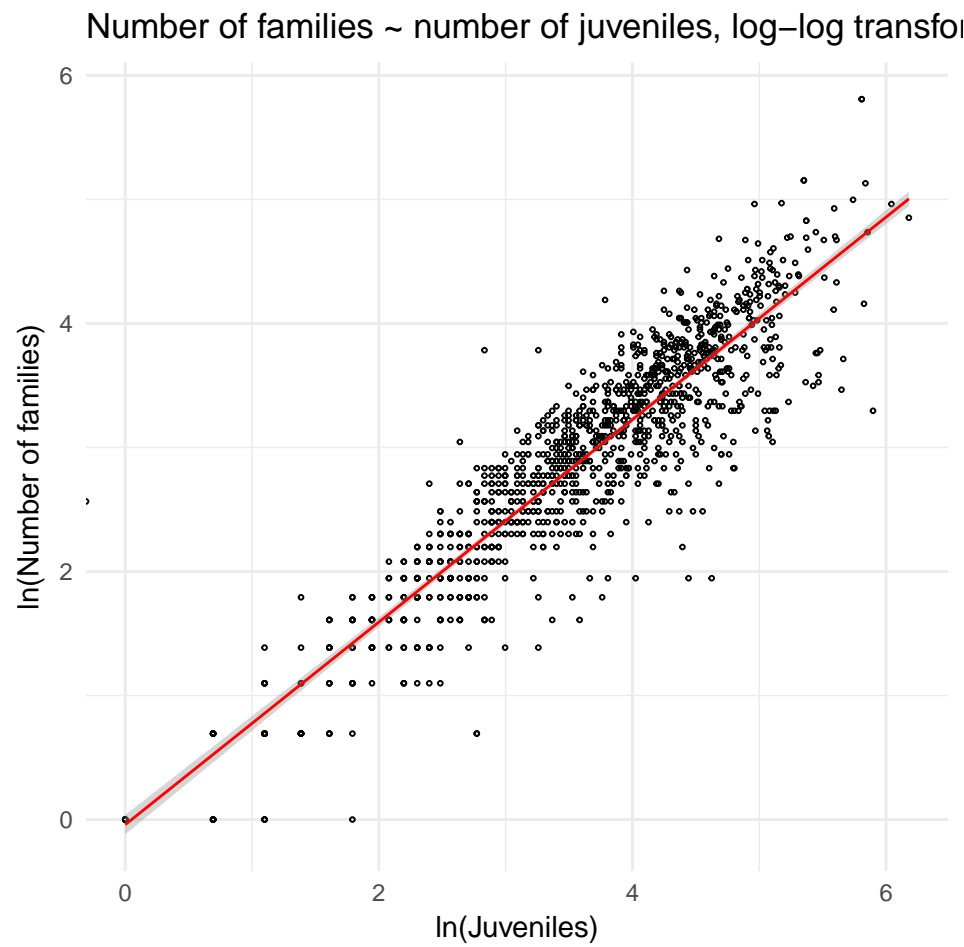
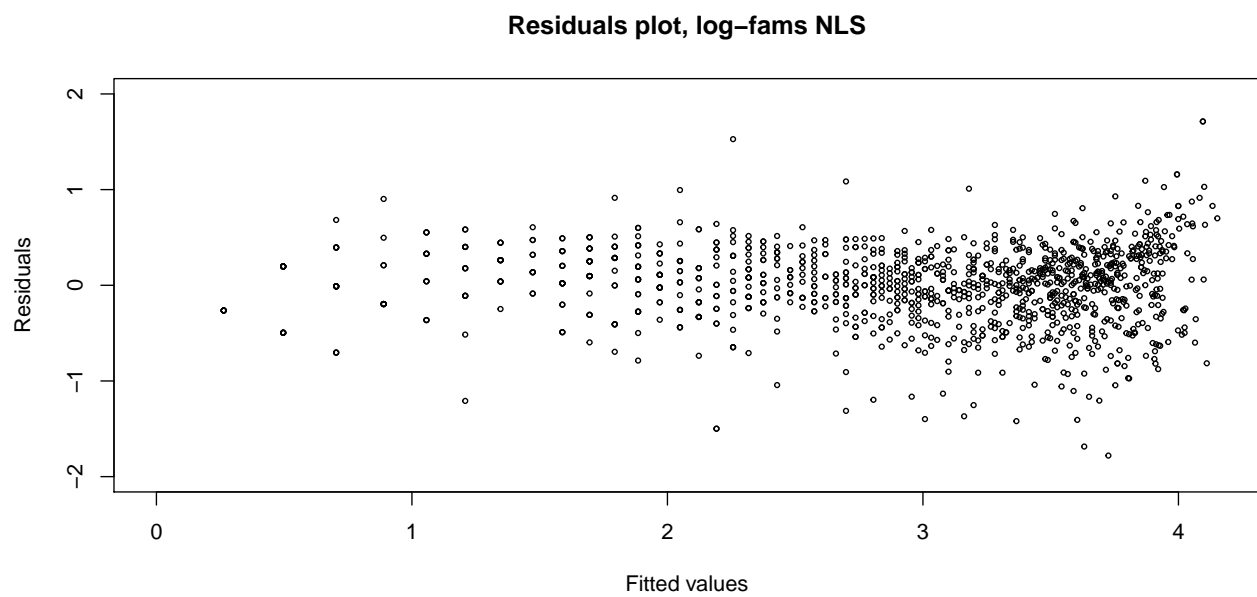
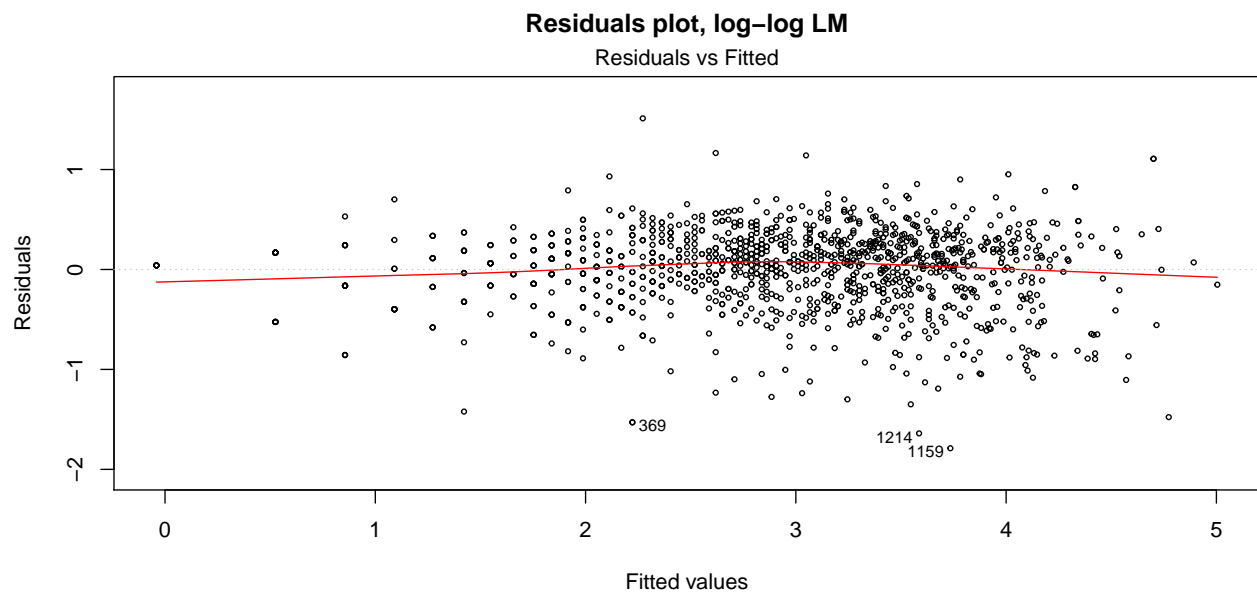
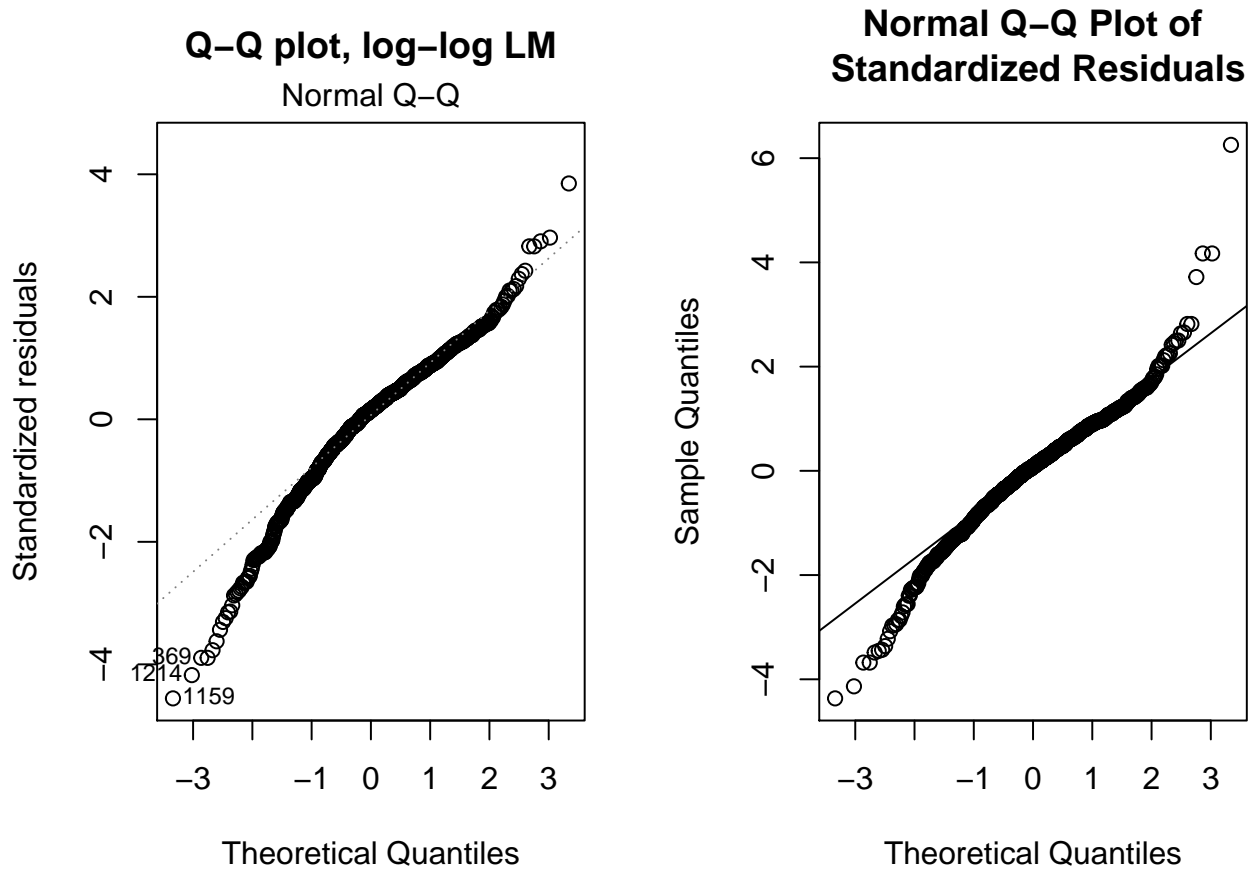


Figure 3: Model fit and data.





## Mean family size ~ number of juveniles

The mean global family size is 1.8663423. The mean family size is the weighted mean of the number of juveniles in each family weighted by the number of families of that size. Ideally, when all juveniles are counted and belong to a family, the mean family size is the number of juveniles by the number of families. In the field, single or unaffiliated juveniles, and imperfect assignment of juveniles to families, contribute to deviations of the true weighted mean family size from the simple mean.

A simple linear model to check whether the simple mean could explain the mean family size found that it could not.

## References

Elder, William H, and Nina L Elder. 1949. "Role of the Family in the Formation of Goose Flocks." *Wilson Bull* 61 (3): 133–40.