Determinants Influencing the Sex Ratio of Turtle Offspring

https://github.com/christalzheng/HuangZheng

Yuechen Huang; Shiqi Zheng

Fall 2023

Contents

Figure 1: A cute baby sea turtle

Contents

List of Tables

List of Figures

1 Rationale and Research Questions

In the majority of species, sex is determined during fertilization; however, reptiles like turtles, alligators, and crocodiles exhibit a unique characteristic where sex determination occurs after fertilization. This process is known as temperature-dependent sex determination (TSD). In TSD, the sex of offspring is influenced by the temperatures experienced during egg incubation in the nest.

Our project is motivated by a keen interest in understanding TSD, particularly in turtles. While TSD was first described in a lizard, the order Testudines has played an important role in advancing scientists' understanding of TSD. Studies on species within Testudines brought TSD to the attention of the broader research community, and the term "temperature-dependent sex determination" also first appeared in a study of a community of North American turtles in 1979 (Yntema, 1976; Bull and Vogt, 1979).

We aim to test the TSD phenomenon by ourself with a focus on the order Testudines and thus choose to investigate a dataset containing plenty sex ratio information for the order Testudines. We will explore the following questions: 1. Does temperature affect the sex ratio of turtles? If so, how does temperature influence the sex ratio of turtles? 2. Is temperature the sole determinant of turtle sex, or are there other important factors?

2 Dataset Information

We use a dataset called Reptilian Offspring Sex and Incubation Environment (ROSIE). It is a dataset published in 2022 and contains over 7,000 individual measurements of offspring sex ratios in the order Testudines. This dataset obtained data by using the Web of Science to search for research published since the discovery of TSD (1966) until 31 December 2020.

We processed the original dataset by omitting all data has been chemical treatment, selecting all TSD type data, all point contains male proportion data, converting date value to date object, and selecting all species information, mean temperature, latitude, longitude, captive, humidity, start date, end date, proportion male, and proportion female as the variables we will explore with.

Variables we used and wrangled for this project are shown in Table 1. More information about the ROSIE Dataset can be found at: https://github.com/calebkrueger/ROSIE/tree/main

Table 1: Summary of Dataset

Used Variables	Details
Species information	Order, Family, Genus, Species
Spatial information	Wild sampling location, or native range of species if captive or location not provided, Provided in average latitude and longitude
Captivity	Eggs from captive or/and wild individuals, 0 - wild and 1 - captive
Time	Time of turtle nesting/egg collection. Provided in start date and end date
Temperature	Mean of actual/recorded incubator temperature in degrees Celsius
Humidity	Relative humidity of incubation chamber between 0 and 1
Sex ratio	Proportion of male

3 Exploratory Analysis

To have a better understanding of our data, we explore the distribution of different turtle species, variations in hatching temperatures and sex ratio across diverse turtle families, and distribution of data location.

3.1 Explore the species in the dataset

In order to explore species information in the dataset, we use the "rotl" package to interact with the Open Tree of Life data (details about this package: https://doi.org/10.1111/2041-210X.12593).

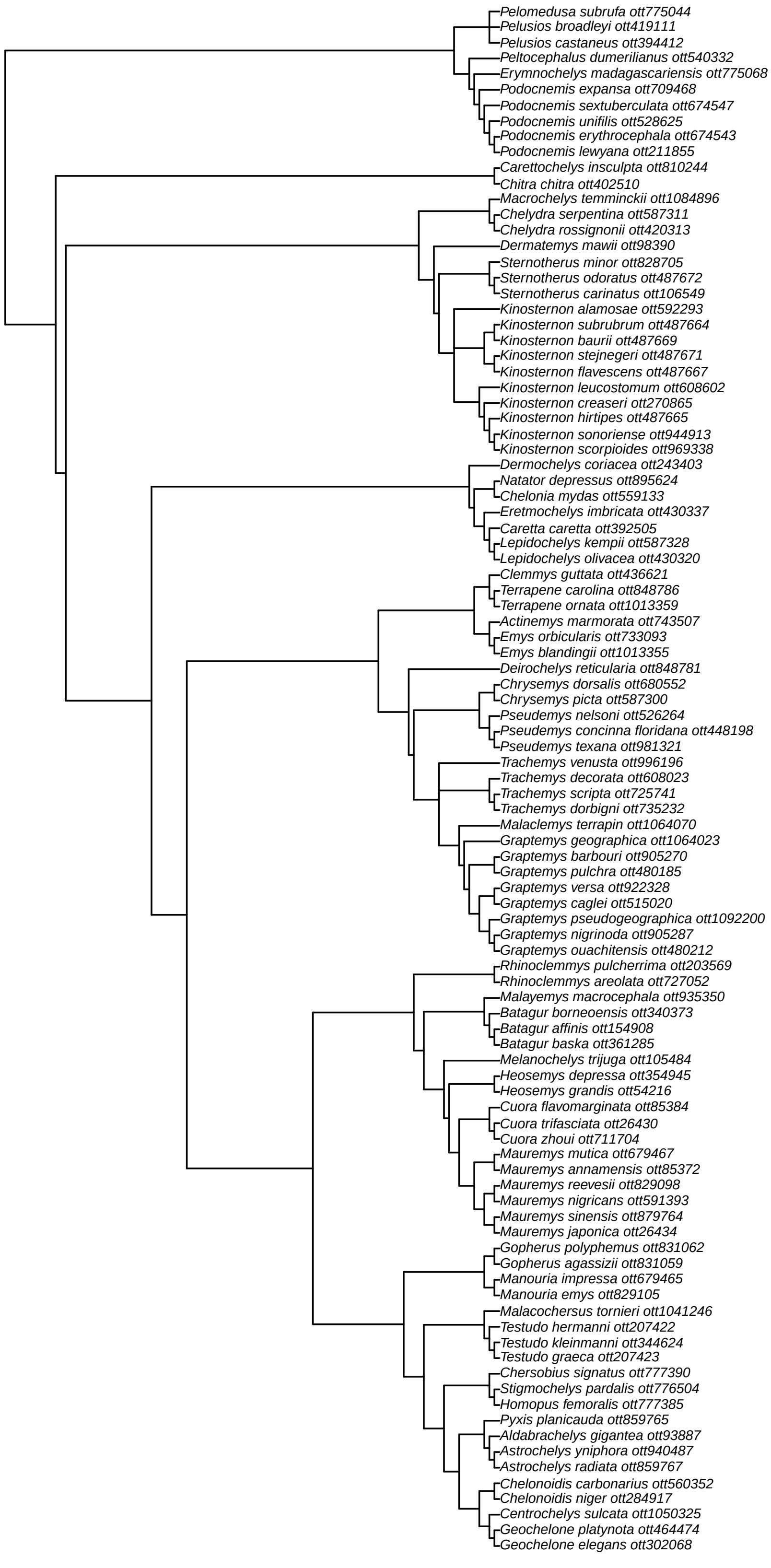

Figure 2: Phylogenetic tree of species in our database

Although the phylogenetic tree clearly shows the relationship between species, it can be too complex and difficult to explore. Therefore, we created a treemap to provide an overview of species information in our data. (code reference for treemap: https://r-graph-gallery.com/236-custom-your-treemap.html).

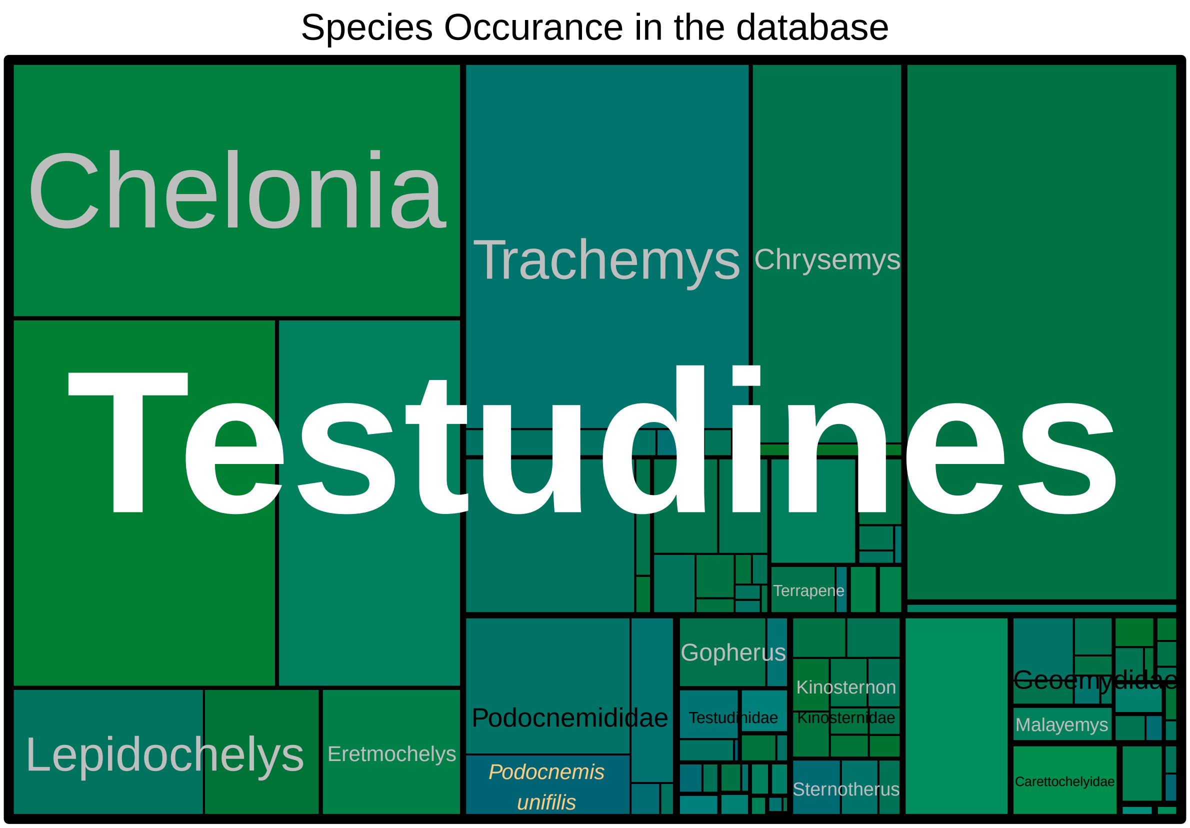

Figure 3: Tree map of species in our database

The interactive tree map can be seen using the link: Interactive Tree Map

```
## \url{../Data/Processed/interactiveTreemap.html}
```

From the tree maps, we can see that all of the species are within Testudines Order (Figure 2 and Figure 3). The most abundant family is Cheloniidae, Emydidae, and Chelydridae. The species we have most in our database are common snapping turtle (Chelydra serpentina) and green sea turtle (Chelonia mydas) with occurrences of 1123 and 881 records.

3.2 Explore hatching temperature among families

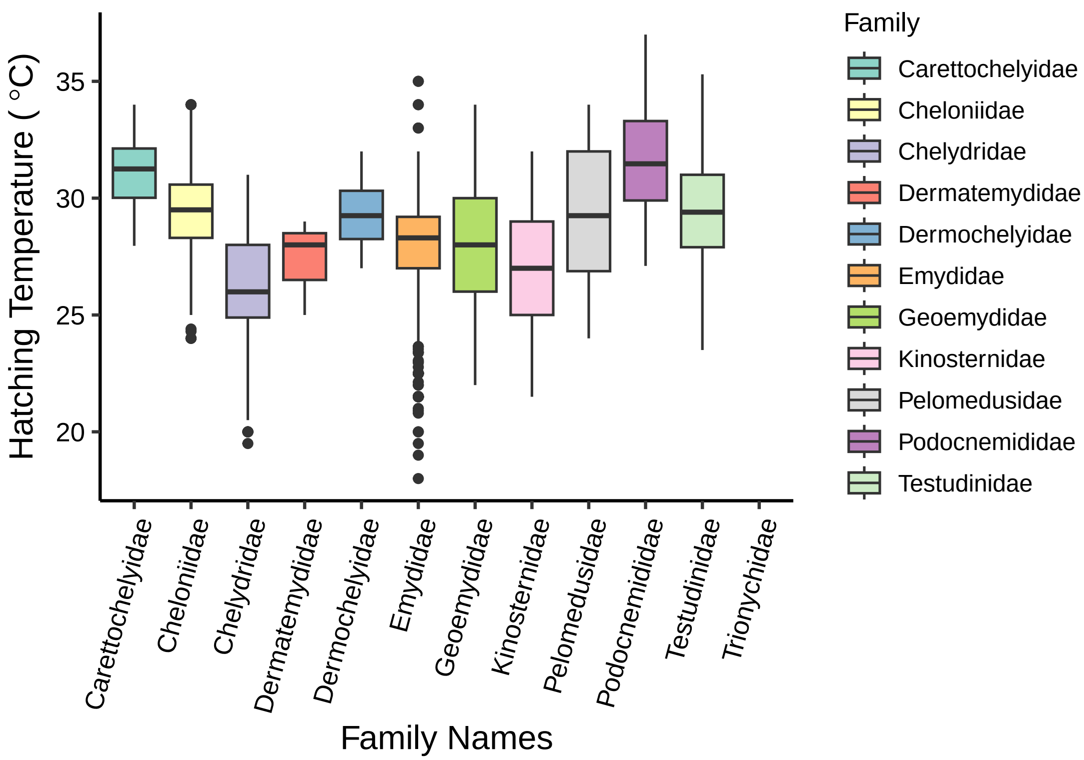

Figure 4: Box plot of temperature among families

The hatching temperature for different turtle families generally falls between 25-35 degree Celsius (Figure 4). However, there is a high variance in the hatching temperature for different individuals or communities of the Emydidae family.

To investigate if there is a significant difference in hatching temperature among families, we use the Shapiro test to check for normality. The null hypothesis is they are not normally distributed.

```
##
##  Shapiro-Wilk normality test
##
## data:  Cleaned_Data$Mean_Temp[Cleaned_Data$Family == "Carettochelyidae"]
## W = 0.95737, p-value = 0.7457
```

```
##
##  Shapiro-Wilk normality test
##
## data:  Cleaned_Data$Mean_Temp[Cleaned_Data$Family == "Cheloniidae"]
## W = 0.99469, p-value = 3.399e-05
```

```
##
##  Shapiro-Wilk normality test
##
## data:  Cleaned_Data$Mean_Temp[Cleaned_Data$Family == "Chelydridae"]
## W = 0.95815, p-value = 6.061e-12
```

```
##
##  Shapiro-Wilk normality test
##
## data:  Cleaned_Data$Mean_Temp[Cleaned_Data$Family == "Dermatemydidae"]
## W = 0.92308, p-value = 0.4633

##
##  Shapiro-Wilk normality test
##
## data:  Cleaned_Data$Mean_Temp[Cleaned_Data$Family == "Dermochelyidae"]
## W = 0.96054, p-value = 0.1544

##
##  Shapiro-Wilk normality test
##
## data:  Cleaned_Data$Mean_Temp[Cleaned_Data$Family == "Emydidae"]
## W = 0.92754, p-value < 2.2e-16

##
##  Shapiro-Wilk normality test
##
## data:  Cleaned_Data$Mean_Temp[Cleaned_Data$Family == "Geoemydidae"]
## W = 0.97459, p-value = 0.01352

##
##  Shapiro-Wilk normality test
##
## data:  Cleaned_Data$Mean_Temp[Cleaned_Data$Family == "Kinosternidae"]
## W = 0.95636, p-value = 3.65e-05

##
##  Shapiro-Wilk normality test
##
## data:  Cleaned_Data$Mean_Temp[Cleaned_Data$Family == "Pelomedusidae"]
## W = 0.93749, p-value = 0.09541

##
##  Shapiro-Wilk normality test
##
## data:  Cleaned_Data$Mean_Temp[Cleaned_Data$Family == "Podocnemididae"]
## W = 0.98295, p-value = 0.07658

##
##  Shapiro-Wilk normality test
##
## data:  Cleaned_Data$Mean_Temp[Cleaned_Data$Family == "Testudinidae"]
## W = 0.98556, p-value = 0.3415
```

The null hypothesis of Cheloniidae, Chelydridae, Dermochelyidae, Emydidae, Geoemydidae, Kinosternidae can be rejected. Thus, Cheloniidae, Chelydridae, Dermochelyidae, Emydidae, Geoemydidae, Kinosternidae are normally distributed.

We can also reject the null hypothesis from the bartlett test ($P = <2.2e-16$) i.e. variances are not equal.

```
##
##  Bartlett test of homogeneity of variances
##
## data:  Cleaned_Data$Mean_Temp and Cleaned_Data$Family
## Bartlett's K-squared = 315.77, df = 10, p-value < 2.2e-16
```

There is a significant difference in hatching temperature among families (F = 184.5, P <2e-16). Families with high variance in hatching temperature from Tukey's HSD test result correspond with the result shown in Figure 4. This variation of temperature among families can be due to genetic variation or the difference in population extinction status.

```
##               Df Sum Sq Mean Sq F value Pr(>F)
## Family        10   7918   791.8   184.5 <2e-16 ***
## Residuals   4090  17557     4.3
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 2578 observations deleted due to missingness
```

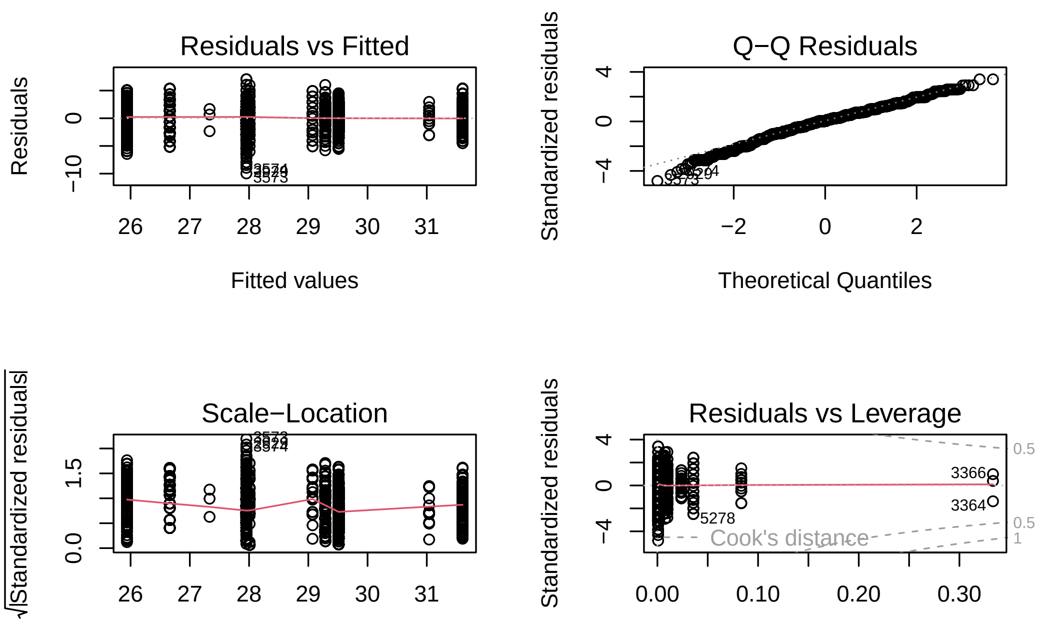

Figure 5: Residual plots for anova test on temperature vs family

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = Mean_Temp ~ Family, data = Cleaned_Data)
##
## $Family
##                                           diff        lwr         upr     p adj
## Cheloniidae-Carettochelyidae    -1.527245395 -3.4610600  0.40656916 0.2795788
```

```
## Chelydridae-Carettochelyidae    -5.099780495 -7.0453328 -3.15422822 0.0000000
## Dermatemydidae-Carettochelyidae -3.705833333 -8.0128601  0.60119339 0.1696238
## Dermochelyidae-Carettochelyidae -1.750595238 -3.9346564  0.43346594 0.2587480
## Emydidae-Carettochelyidae       -3.084400243 -5.0189785 -1.14982199 0.0000160
## Geoemydidae-Carettochelyidae    -3.030294486 -5.0414736 -1.01911540 0.0000674
## Kinosternidae-Carettochelyidae  -4.377763158 -6.3703628 -2.38516349 0.0000000
## Pelomedusidae-Carettochelyidae  -1.967738095 -4.2699407  0.33446453 0.1771400
## Podocnemididae-Carettochelyidae  0.564762411 -1.4416894  2.57121418 0.9981480
## Testudinidae-Carettochelyidae   -1.753523102 -3.7908987  0.28385249 0.1692840
## Chelydridae-Cheloniidae         -3.572535100 -3.8959864 -3.24908383 0.0000000
## Dermatemydidae-Cheloniidae      -2.178587938 -6.0347423  1.67756640 0.7687402
## Dermochelyidae-Cheloniidae      -0.223349843 -1.2671747  0.82047499 0.9998343
## Emydidae-Cheloniidae            -1.557154848 -1.8062333 -1.30807637 0.0000000
## Geoemydidae-Cheloniidae         -1.503049091 -2.1066122 -0.89948601 0.0000000
## Kinosternidae-Cheloniidae       -2.850517763 -3.3889419 -2.31209367 0.0000000
## Pelomedusidae-Cheloniidae       -0.440492700 -1.7131215  0.83213607 0.9900672
## Podocnemididae-Cheloniidae       2.092007806  1.5043891  2.67962651 0.0000000
## Testudinidae-Cheloniidae        -0.226277707 -0.9120957  0.45954032 0.9931761
## Dermatemydidae-Chelydridae       1.393947161 -2.4681069  5.25600118 0.9863144
## Dermochelyidae-Chelydridae       3.349185257  2.2837721  4.41459842 0.0000000
## Emydidae-Chelydridae             2.015380251  1.6873939  2.34336655 0.0000000
## Geoemydidae-Chelydridae          2.069486008  1.4293116  2.70966045 0.0000000
## Kinosternidae-Chelydridae        0.722017337  0.1428495  1.30118521 0.0029539
## Pelomedusidae-Chelydridae        3.132042399  1.8416475  4.42243726 0.0000000
## Podocnemididae-Chelydridae       5.664542906  5.0393784  6.28970741 0.0000000
## Testudinidae-Chelydridae         3.346257392  2.6280086  4.06450616 0.0000000
## Dermochelyidae-Dermatemydidae    1.955238095 -2.0322938  5.94277001 0.8913008
## Emydidae-Dermatemydidae          0.621433090 -3.2351043  4.47797047 0.9999881
## Geoemydidae-Dermatemydidae       0.675538847 -3.2199879  4.57106562 0.9999763
## Kinosternidae-Dermatemydidae    -0.671929825 -4.5578970  3.21403736 0.9999769
## Pelomedusidae-Dermatemydidae     1.738095238 -2.3153508  5.79154124 0.9531534
## Podocnemididae-Dermatemydidae    4.270595745  0.3775075  8.16368401 0.0180935
## Testudinidae-Dermatemydidae      1.952310231 -1.9568056  5.86142610 0.8791349
## Emydidae-Dermochelyidae         -1.333805005 -2.3790440 -0.28856600 0.0020021
## Geoemydidae-Dermochelyidae      -1.279699248 -2.4607043 -0.09869422 0.0209831
## Kinosternidae-Dermochelyidae    -2.627167920 -3.7762480 -1.47808786 0.0000000
## Pelomedusidae-Dermochelyidae    -0.217142857 -1.8450459  1.41076023 0.9999981
## Podocnemididae-Dermochelyidae    2.315357649  1.1424211  3.48829423 0.0000000
## Testudinidae-Dermochelyidae     -0.002927864 -1.2280119  1.22215615 1.0000000
## Geoemydidae-Emydidae             0.054105757 -0.5518998  0.66011127 1.0000000
## Kinosternidae-Emydidae          -1.293362915 -1.8345235 -0.75220232 0.0000000
## Pelomedusidae-Emydidae           1.116662148 -0.1571268  2.39045109 0.1494776
## Podocnemididae-Emydidae          3.649162655  3.0590355  4.23928978 0.0000000
## Testudinidae-Emydidae            1.330877141  0.6429086  2.01884564 0.0000001
## Kinosternidae-Geoemydidae       -1.347468672 -2.1188979 -0.57603942 0.0000011
## Pelomedusidae-Geoemydidae        1.062556391 -0.3248104  2.44992317 0.3247262
## Podocnemididae-Geoemydidae       3.595056898  2.7885221  4.40159166 0.0000000
## Testudinidae-Geoemydidae         1.276771384  0.3961188  2.15742396 0.0001644
## Pelomedusidae-Kinosternidae      2.410025063  1.0497315  3.77031866 0.0000007
## Podocnemididae-Kinosternidae     4.942525569  4.1835062  5.70154496 0.0000000
## Testudinidae-Kinosternidae       2.624240056  1.7868864  3.46159367 0.0000000
## Podocnemididae-Pelomedusidae     2.532500507  1.1519956  3.91300546 0.0000002
## Testudinidae-Pelomedusidae       0.214214993 -1.2108620  1.63929202 0.9999939
## Testudinidae-Podocnemididae     -2.318285514 -3.1880880 -1.44848306 0.0000000
```

3.3 Explore sex ratio among families

For this project, we define sex ratio as the proportion of males in the population. We abandon the male/female sex ratio definition since it can be rendered meaningless when the number of females is zero.

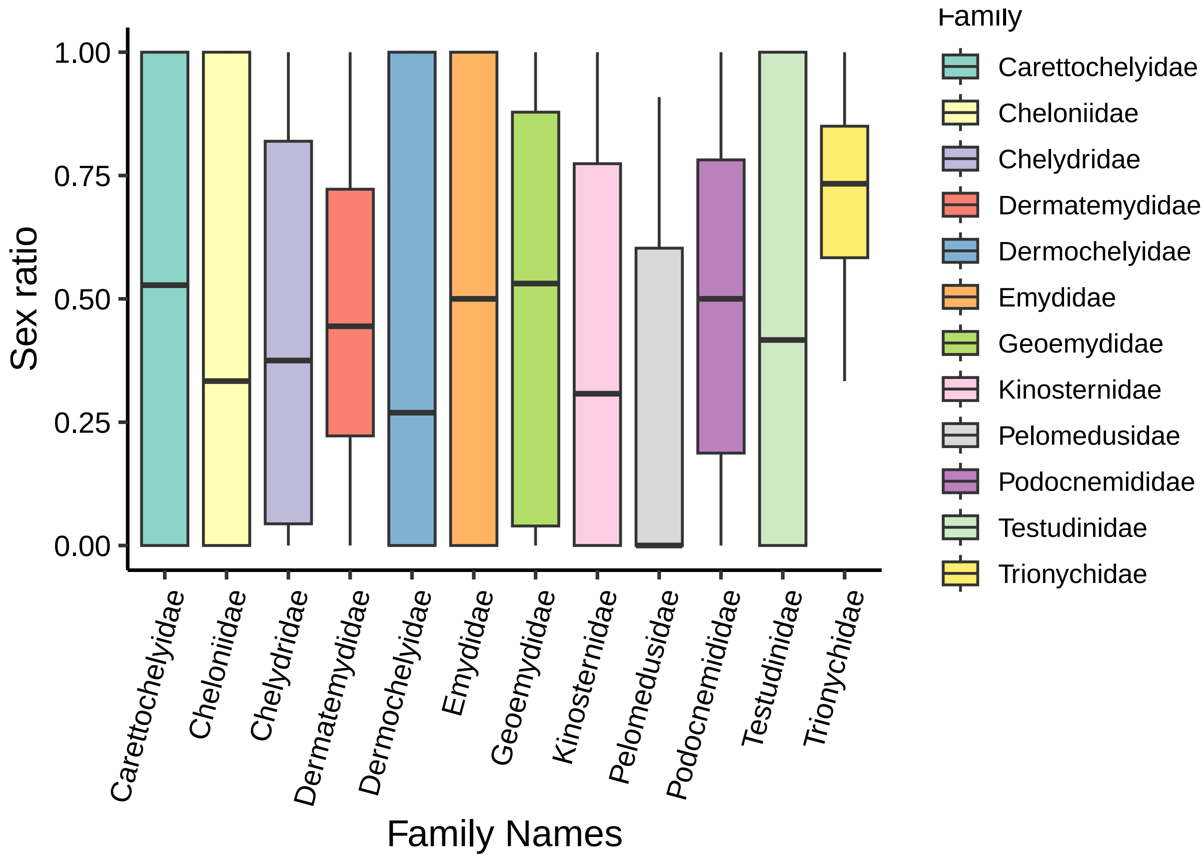

Figure 6: Box plot of sex ratio among families

The sex ratio seems to vary among different families (Figure 6). According to the figure 6, Pelomedusidae, Cheloniidae, Chelydridae, Dermochelyidae, Kinosternidae have a mean average sex ratio that is lower than 0.5 (i.e. more female). Trionychidae has a sex ratio that is clearly greater than 0.5 (i.e. more male). These families may suffer from species extinction because of the unbalanced population. All the families have very high variance. This may be because of the shortage of data (so that extreme values can have more impact on the total data). This indicates the necessity of having more conservation effort in data collection or population management.

We want to explore whether there are significant differences in sex ratio between different families. We use One-way ANOVA to analyze the data. First, we test the normality:

```
##
##  Shapiro-Wilk normality test
##
## data:  Cleaned_Data$Proportion_Male[Cleaned_Data$Family == "Carettochelyidae"]
## W = 0.73956, p-value = 3.734e-09

##
##  Shapiro-Wilk normality test
```

```
##
## data:  Cleaned_Data$Proportion_Male[Cleaned_Data$Family == "Cheloniidae"]
## W = 0.78602, p-value < 2.2e-16

##
##  Shapiro-Wilk normality test
##
## data:  Cleaned_Data$Proportion_Male[Cleaned_Data$Family == "Chelydridae"]
## W = 0.86642, p-value < 2.2e-16

##
##  Shapiro-Wilk normality test
##
## data:  Cleaned_Data$Proportion_Male[Cleaned_Data$Family == "Dermatemydidae"]
## W = 0.9959, p-value = 0.8777

##
##  Shapiro-Wilk normality test
##
## data:  Cleaned_Data$Proportion_Male[Cleaned_Data$Family == "Dermochelyidae"]
## W = 0.73932, p-value = 8.553e-16

##
##  Shapiro-Wilk normality test
##
## data:  Cleaned_Data$Proportion_Male[Cleaned_Data$Family == "Emydidae"]
## W = 0.81127, p-value < 2.2e-16

##
##  Shapiro-Wilk normality test
##
## data:  Cleaned_Data$Proportion_Male[Cleaned_Data$Family == "Geoemydidae"]
## W = 0.859, p-value = 2.944e-11

##
##  Shapiro-Wilk normality test
##
## data:  Cleaned_Data$Proportion_Male[Cleaned_Data$Family == "Kinosternidae"]
## W = 0.85597, p-value = 1.007e-11

##
##  Shapiro-Wilk normality test
##
## data:  Cleaned_Data$Proportion_Male[Cleaned_Data$Family == "Pelomedusidae"]
## W = 0.6819, p-value = 1.597e-06

##
##  Shapiro-Wilk normality test
##
## data:  Cleaned_Data$Proportion_Male[Cleaned_Data$Family == "Podocnemididae"]
## W = 0.92871, p-value = 2.16e-11
```

```
##
##  Shapiro-Wilk normality test
##
## data:  Cleaned_Data$Proportion_Male[Cleaned_Data$Family == "Testudinidae"]
## W = 0.79178, p-value = 2.012e-14
```

```
##
##  Shapiro-Wilk normality test
##
## data:  Cleaned_Data$Proportion_Male[Cleaned_Data$Family == "Trionychidae"]
## W = 0.97956, p-value = 0.8994
```

Apart from Dermatemydidae and Trionychida, all data follows a normal distribution. We then check equality of variance:

```
##
##  Bartlett test of homogeneity of variances
##
## data:  Cleaned_Data$Proportion_Male and Cleaned_Data$Family
## Bartlett's K-squared = 56.623, df = 11, p-value = 3.903e-08
```

We can reject the null hypothesis (P = 3.903e-08) i.e. variances are not equal.

```
##               Df Sum Sq Mean Sq F value   Pr(>F)
## Family        11    9.5  0.8638   5.119 4.85e-08 ***
## Residuals   6667 1125.0  0.1687
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = Proportion_Male ~ Family, data = Cleaned_Data)
##
## $Family
##                                          diff         lwr         upr
## Cheloniidae-Carettochelyidae    -0.0729091376 -0.24548861  0.099670338
## Chelydridae-Carettochelyidae    -0.0659573084 -0.24102751  0.109112891
## Dermatemydidae-Carettochelyidae -0.0252931304 -0.81915400  0.768567736
## Dermochelyidae-Carettochelyidae -0.0619818009 -0.26202293  0.138059329
## Emydidae-Carettochelyidae        0.0034263455 -0.16997343  0.176826123
## Geoemydidae-Carettochelyidae    -0.0162922579 -0.21649989  0.183915379
## Kinosternidae-Carettochelyidae  -0.0962752550 -0.29520148  0.102650973
## Pelomedusidae-Carettochelyidae  -0.2725437025 -0.57831068  0.033223276
## Podocnemididae-Carettochelyidae -0.0243540152 -0.21036957  0.161661542
## Testudinidae-Carettochelyidae   -0.0217212721 -0.22049511  0.177052566
## Trionychidae-Carettochelyidae    0.1932253882 -0.49954659  0.885997370
## Chelydridae-Cheloniidae          0.0069518292 -0.04058290  0.054486560
## Dermatemydidae-Cheloniidae       0.0476160072 -0.72815783  0.823389848
## Dermochelyidae-Cheloniidae       0.0109273366 -0.09689859  0.118753260
## Emydidae-Cheloniidae             0.0763354831  0.03537869  0.117292277
## Geoemydidae-Cheloniidae          0.0566168797 -0.05151764  0.164751396
```

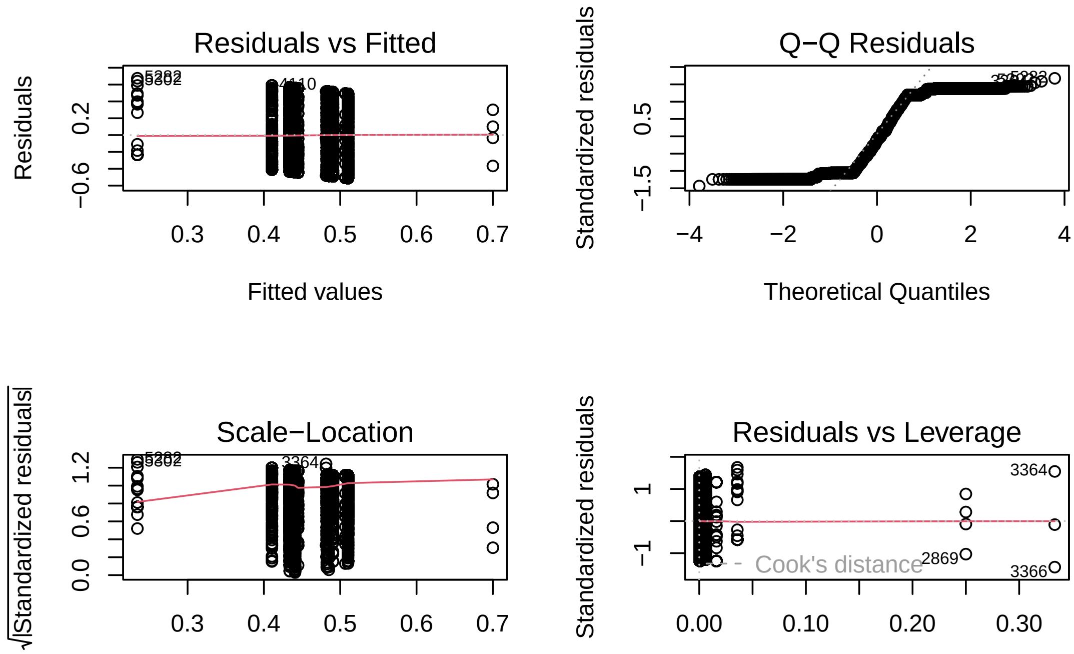

Figure 7: Residual plots for anova test on sex ratio vs family

```
## Kinosternidae-Cheloniidae       -0.0233661175 -0.12910930  0.082377063
## Pelomedusidae-Cheloniidae       -0.1996345649 -0.45478829  0.055519156
## Podocnemididae-Cheloniidae       0.0485551223 -0.03025962  0.127369864
## Testudinidae-Cheloniidae         0.0511878655 -0.05426836  0.156644086
## Trionychidae-Cheloniidae         0.2661345258 -0.40583496  0.938104016
## Dermatemydidae-Chelydridae       0.0406641780 -0.73566755  0.816995905
## Dermochelyidae-Chelydridae       0.0039755075 -0.10779358  0.115744596
## Emydidae-Chelydridae             0.0693836540  0.01895193  0.119815374
## Geoemydidae-Chelydridae          0.0496650506 -0.06240177  0.161731875
## Kinosternidae-Chelydridae       -0.0303179466 -0.14007914  0.079443247
## Pelomedusidae-Chelydridae       -0.2065863941 -0.46343133  0.050258541
## Podocnemididae-Chelydridae       0.0416032932 -0.04252555  0.125732140
## Testudinidae-Chelydridae         0.0442360364 -0.06524873  0.153720802
## Trionychidae-Chelydridae         0.2591826967 -0.41343078  0.931796177
## Dermochelyidae-Dermatemydidae   -0.0366886705 -0.81902992  0.745652579
## Emydidae-Dermatemydidae          0.0287194760 -0.74723726  0.804676214
## Geoemydidae-Dermatemydidae       0.0090008726 -0.77338297  0.791384713
## Kinosternidae-Dermatemydidae    -0.0709821246 -0.85303904  0.711074792
## Pelomedusidae-Dermatemydidae    -0.2472505721 -1.06305375  0.568552604
## Podocnemididae-Dermatemydidae    0.0009391152 -0.77793389  0.779812124
## Testudinidae-Dermatemydidae      0.0035718584 -0.77844631  0.785590026
## Trionychidae-Dermatemydidae      0.2185185187 -0.80713951  1.244176549
## Emydidae-Dermochelyidae          0.0654081465 -0.04372589  0.174542181
## Geoemydidae-Dermochelyidae       0.0456895431 -0.10238426  0.193763349
## Kinosternidae-Dermochelyidae    -0.0342934541 -0.18063005  0.112043140
## Pelomedusidae-Dermochelyidae    -0.2105619016 -0.48503631  0.063912505
```

```
## Podocnemididae-Dermochelyidae    0.0376277857 -0.09060849  0.165864060
## Testudinidae-Dermochelyidae      0.0402605289 -0.10586884  0.186389900
## Trionychidae-Dermochelyidae      0.2552071892 -0.42433367  0.934748044
## Geoemydidae-Emydidae            -0.0197186034 -0.12915754  0.089720336
## Kinosternidae-Emydidae          -0.0997016006 -0.20677834  0.007375138
## Pelomedusidae-Emydidae          -0.2759700481 -0.53167931 -0.020260782
## Podocnemididae-Emydidae         -0.0277803608 -0.10837547  0.052814746
## Testudinidae-Emydidae           -0.0251476176 -0.13194098  0.081645745
## Trionychidae-Emydidae            0.1897990427 -0.48238159  0.861979675
## Kinosternidae-Geoemydidae       -0.0799829972 -0.22654712  0.066581128
## Pelomedusidae-Geoemydidae       -0.2562514446 -0.53084723  0.018344338
## Podocnemididae-Geoemydidae      -0.0080617574 -0.13655762  0.120434102
## Testudinidae-Geoemydidae        -0.0054290142 -0.15178624  0.140928210
## Trionychidae-Geoemydidae         0.2095176461 -0.47007224  0.889107535
## Pelomedusidae-Kinosternidae     -0.1762684475 -0.44993136  0.097394466
## Podocnemididae-Kinosternidae     0.0719212398 -0.05456881  0.198411289
## Testudinidae-Kinosternidae       0.0745539830 -0.07004541  0.219153373
## Trionychidae-Kinosternidae       0.2895006433 -0.38971285  0.968714133
## Podocnemididae-Pelomedusidae     0.2481896873 -0.01623711  0.512616489
## Testudinidae-Pelomedusidae       0.2508224305 -0.02272973  0.524374591
## Trionychidae-Pelomedusidae       0.4657690908 -0.25204200  1.183580184
## Testudinidae-Podocnemididae      0.0026327432 -0.12361751  0.128882998
## Trionychidae-Podocnemididae      0.2175794035 -0.45796564  0.893124448
## Trionychidae-Testudinidae        0.2149466603 -0.46422221  0.894115534
##                                     p adj
## Cheloniidae-Carettochelyidae    0.9672388
## Chelydridae-Carettochelyidae    0.9864975
## Dermatemydidae-Carettochelyidae 1.0000000
## Dermochelyidae-Carettochelyidae 0.9974500
## Emydidae-Carettochelyidae       1.0000000
## Geoemydidae-Carettochelyidae    1.0000000
## Kinosternidae-Carettochelyidae  0.9157838
## Pelomedusidae-Carettochelyidae  0.1360240
## Podocnemididae-Carettochelyidae 0.9999995
## Testudinidae-Carettochelyidae   0.9999999
## Trionychidae-Carettochelyidae   0.9990242
## Chelydridae-Cheloniidae         0.9999985
## Dermatemydidae-Cheloniidae      1.0000000
## Dermochelyidae-Cheloniidae      1.0000000
## Emydidae-Cheloniidae            0.0000001
## Geoemydidae-Cheloniidae         0.8632670
## Kinosternidae-Cheloniidae       0.9998966
## Pelomedusidae-Cheloniidae       0.3042600
## Podocnemididae-Cheloniidae      0.6839371
## Testudinidae-Cheloniidae        0.9141764
## Trionychidae-Cheloniidae        0.9799379
## Dermatemydidae-Chelydridae      1.0000000
## Dermochelyidae-Chelydridae      1.0000000
## Emydidae-Chelydridae            0.0004330
## Geoemydidae-Chelydridae         0.9536613
## Kinosternidae-Chelydridae       0.9991088
## Pelomedusidae-Chelydridae       0.2631715
## Podocnemididae-Chelydridae      0.9033616
## Testudinidae-Chelydridae        0.9765910
```

```
## Trionychidae-Chelydridae        0.9838295
## Dermochelyidae-Dermatemydidae   1.0000000
## Emydidae-Dermatemydidae         1.0000000
## Geoemydidae-Dermatemydidae      1.0000000
## Kinosternidae-Dermatemydidae    1.0000000
## Pelomedusidae-Dermatemydidae    0.9979091
## Podocnemididae-Dermatemydidae   1.0000000
## Testudinidae-Dermatemydidae     1.0000000
## Trionychidae-Dermatemydidae     0.9999281
## Emydidae-Dermochelyidae         0.7211465
## Geoemydidae-Dermochelyidae      0.9975432
## Kinosternidae-Dermochelyidae    0.9998154
## Pelomedusidae-Dermochelyidae    0.3348723
## Podocnemididae-Dermochelyidae   0.9984417
## Testudinidae-Dermochelyidae     0.9991298
## Trionychidae-Dermochelyidae     0.9868376
## Geoemydidae-Emydidae            0.9999869
## Kinosternidae-Emydidae          0.0963845
## Pelomedusidae-Emydidae          0.0215075
## Podocnemididae-Emydidae         0.9935362
## Testudinidae-Emydidae           0.9998065
## Trionychidae-Emydidae           0.9989062
## Kinosternidae-Geoemydidae       0.8270159
## Pelomedusidae-Geoemydidae       0.0946033
## Podocnemididae-Geoemydidae      1.0000000
## Testudinidae-Geoemydidae        1.0000000
## Trionychidae-Geoemydidae        0.9975616
## Pelomedusidae-Kinosternidae     0.6187033
## Podocnemididae-Kinosternidae    0.7847244
## Testudinidae-Kinosternidae      0.8751467
## Trionychidae-Kinosternidae      0.9650261
## Podocnemididae-Pelomedusidae    0.0900855
## Testudinidae-Pelomedusidae      0.1093354
## Trionychidae-Pelomedusidae      0.6073720
## Testudinidae-Podocnemididae     1.0000000
## Trionychidae-Podocnemididae     0.9964050
## Trionychidae-Testudinidae       0.9969202
```

From the results, we can see that there is a significant difference in sex ratio among families. Emydidae-Cheloniidae, Emydidae-Chelydridae and Pelomedusidae-Emydidae are the groups that show significant difference ($p < 0.05$). The reason could be that the three families are more genetically different from each other than other families.

3.4 Explore data position

The location of wild sampling points, or the native range of species if captive, is plotted below as a map to show the spatial distribution of data. The world continent data is downloaded from ArcGIS database.

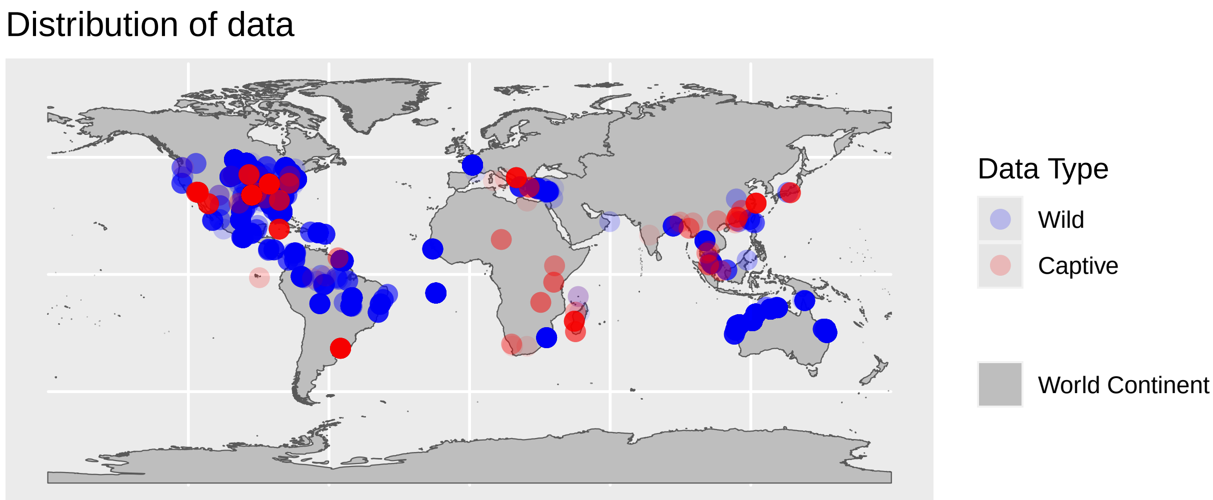

Figure 8: Spatial distribution of data

4 Analysis

4.1 Question 1: Does temperature affect the sex ratio of turtles?

To answer this question, we first want to visually inspect the influence of temperature on the sex ratio. Here, we created a scatter plot with a smooth line showing the trend. We fitted a linear regression line in orange to the plot.

```
## `geom_smooth()` using formula = 'y ~ x'
```

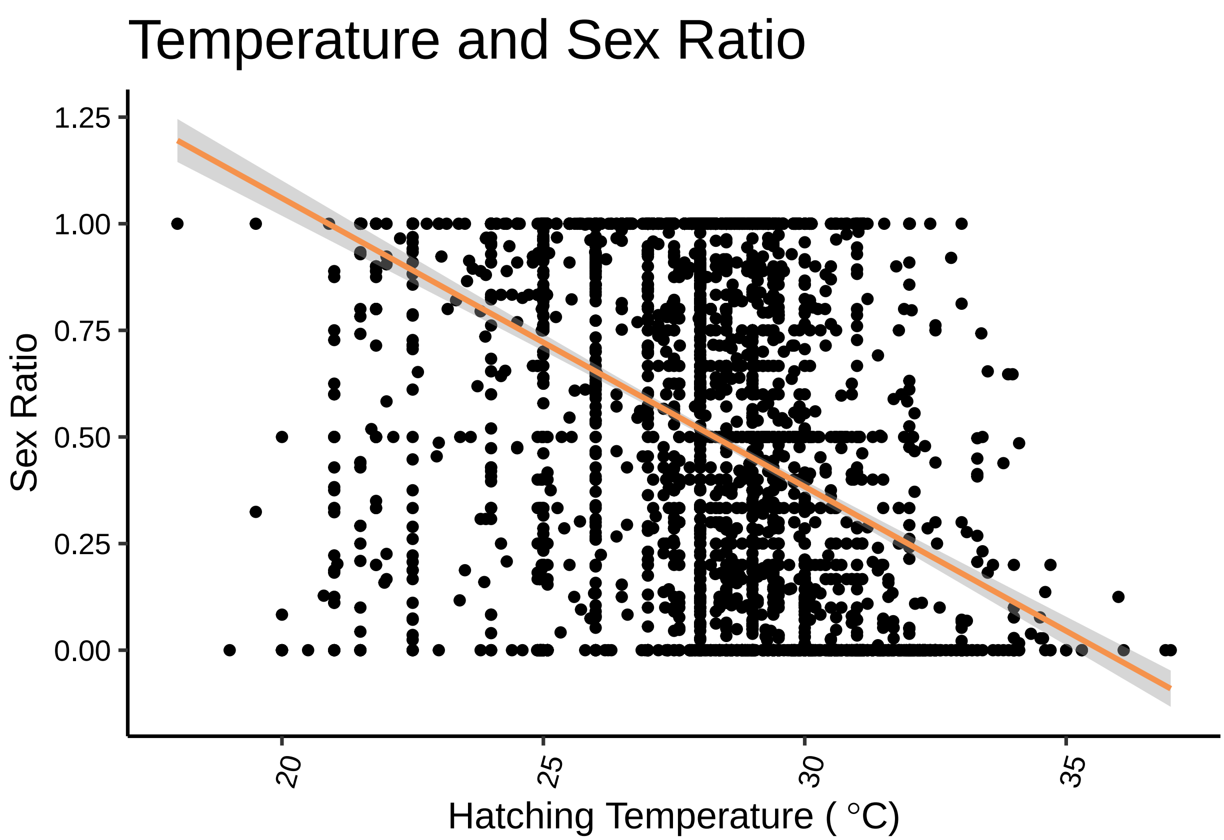

Figure 9: Scatter plot of temperature and sex ratio

The linear trend can be seen on the scatter plot. As temperature increases, the sex ratio decreases. This trend corresponds to our expectation as higher temperature may lead to more female population. Yet, a more sophisticated linear regression analysis is still needed to be assured.

Null Hypothesis: There is no effect of temperature on the sex ratio of turtles. Alternative Hypothesis: There is an effect of temperature on the sex ratio of turtles.

```
##
## Call:
## lm(formula = Proportion_Male ~ Mean_Temp, data = Data_omit)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
```

```
## -1.12760 -0.34413 -0.01868  0.36058  0.81960
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.413091   0.068801   35.07   <2e-16 ***
## Mean_Temp   -0.067657   0.002416  -28.01   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3856 on 4099 degrees of freedom
## Multiple R-squared:  0.1606, Adjusted R-squared:  0.1604
## F-statistic: 784.4 on 1 and 4099 DF,  p-value: < 2.2e-16
```

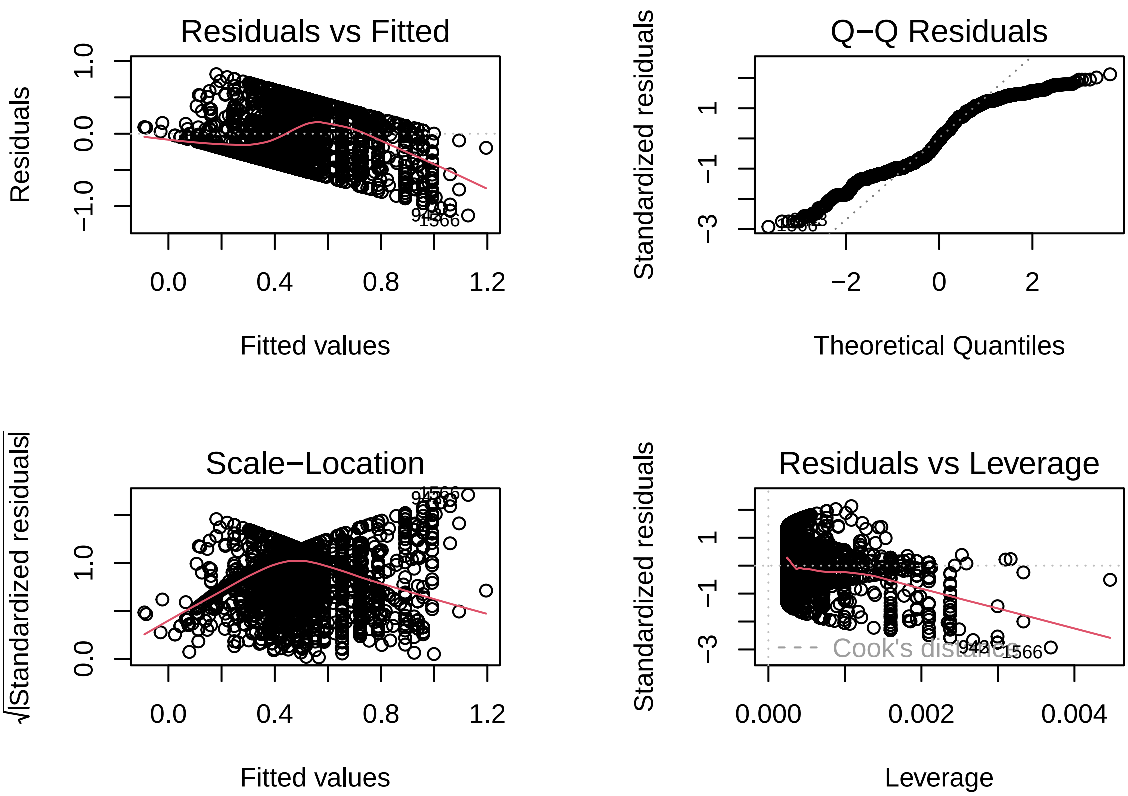

Figure 10: Linear regression of temperature and sex ratio

According to the model, we find that R-squared value is 0.1606 which is very low meaning that only a small proportion of data can be explained by this linear regression model. However, we still assured that temperature has a great impact on sex ratio after rejecting the null hypothesis (P < 2e-16).

4.1.1 How does temperature influence the sex ratio of turtles?

After some literature reviews, we learned the relationship of temperature and sex ratio shown in the table 2. When temperature is between 28 to 32 °C, the sex ratio (male proportion in our study) is approximately 0.5. Temperature below 28 °C results in a higher proportion of male while temperature above 32 °C leads to more female. We can see from our Figure 9 that from 28 - 32 degree Celsius, the number of points lay around sex ratio = 0.5 increases.

Table 2: Temperature-Sex Ratio Relationship

Temperature	Sex Ratio
> 32°C	Female
28-32 °C	Male:Female ~ 50:50
< 28°C	Male

4.1.2 Does the impact of temperature on sex ratio differentiate among families?

We then use a bubble plot to visualize the impact of temperature on sex ratio among species.

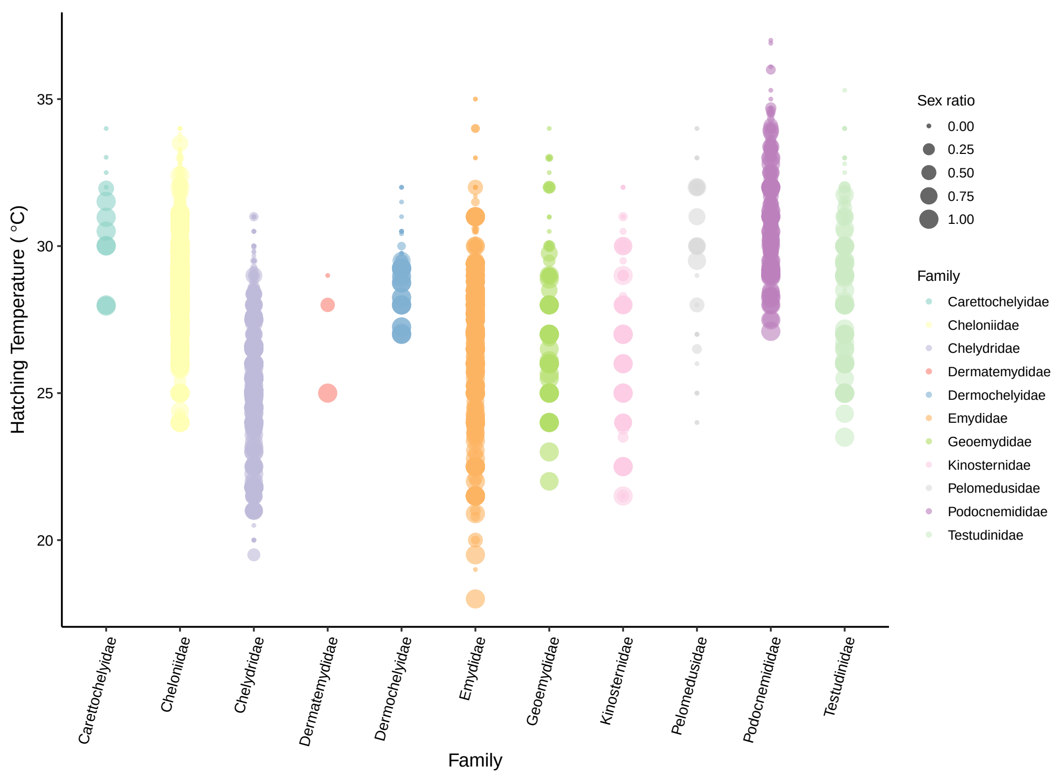

Figure 11: The impact of temperature on sex ratio among families

The impact of temperature on sex ratio exists among most of the families including Carettochelyidae, Cheloniidae, Dermatemydidae, Geoemydidae, Kinosternidae, Podocnemididae and Testudinidae (Figure 11). We can see from the graph that as the temperature goes higher, the sex ratio of these families goes lower. However, Chelydridae, Dermochelyidae, Emydidae and Pelomedusidae have a trend that when temperature is lower than a certain value, sex ratio becomes higher as the temperature goes higher. Yet, when temperature is higher than a certain value, the trend reverses.

4.2 Question 2: Are there other important factors influence sex ratio of turtles?

To answer this question, we explore the impacts of temporal and spatial variation on the sex ratio of turtles.

4.2.1 Temporal impacts

At Year Level Null Hypothesis: There is no effect of year on the sex ratio of turtles. Alternative Hypothesis: There is an effect of year on the sex ratio of turtles.

Based on the linear model, we rejects the null hypothesis with a p value of 2.873e-16. This p-value also shows this model is statistically significant and a meaningful regression. But according to the R^2 value, only 1.495% of the total variance in sex ratio is explained by changes in year. The sex ratio is predicted to increase 0.0041539 with 1 year change. The results of model are clearly showed in Figure 12. The orange line shows a overall slightly increasing trend between year and sex ratio, indicating that as time passed by there are more male turtles. This is contradicts with our intuition that time increases will cause temperature increases and thus will lead to more female turtles. But because temperature trends might not be strictly linear or may be influenced by other factors in long term scenario, and the very noisy scatter plot and low R^2 value both show that this pattern only explain limited part of data, the linear model results of year and sex ratio is still statistically significant and a meaningful regression.

Below are the results of the linear model and a scatter plot of year and sex ratio.

```
##
## Call:
## lm(formula = Proportion_Male ~ Year, data = Cleaned_Data_omit)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.52675 -0.42290 -0.03557  0.48156  0.63525
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -7.8599014  1.0128760   -7.76 1.05e-14 ***
## Year         0.0041539  0.0005059    8.21 2.87e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.412 on 4376 degrees of freedom
## Multiple R-squared:  0.01517,    Adjusted R-squared:  0.01495
## F-statistic: 67.41 on 1 and 4376 DF,  p-value: 2.873e-16

## `geom_smooth()` using formula = 'y ~ x'
```

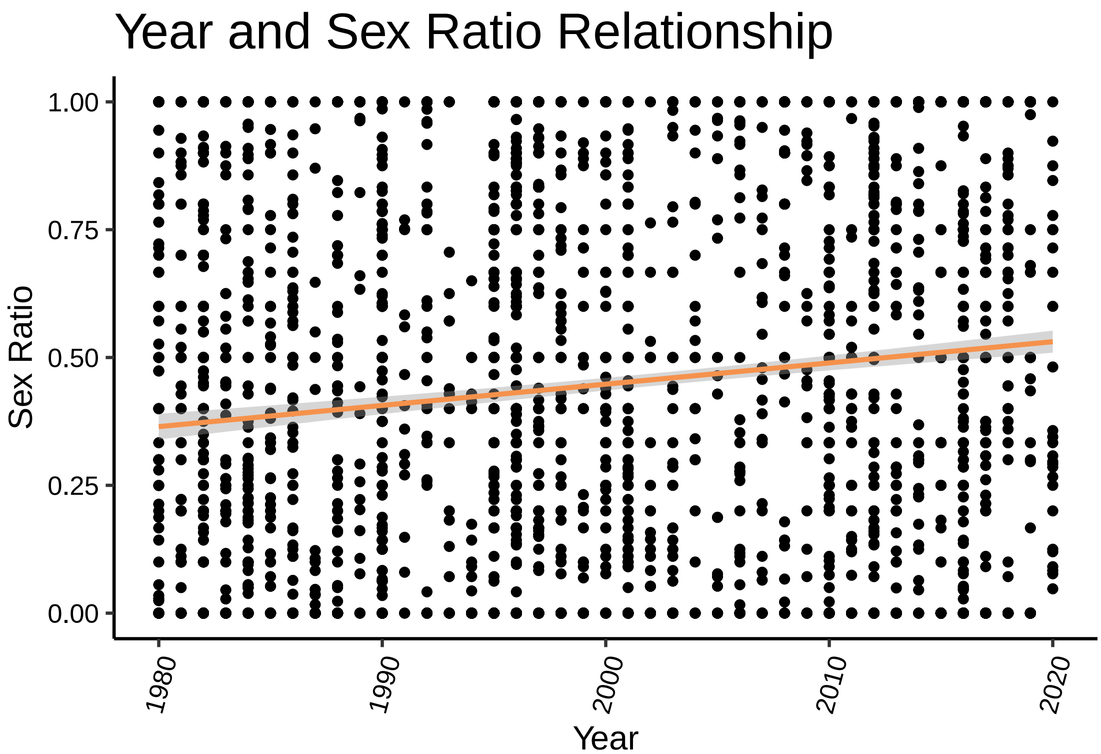

Figure 12: Year and sex ratio relationship

At Month Level Null Hypothesis: There is no effect of month on the sex ratio of turtles. Alternative Hypothesis: There is an effect of month on the sex ratio of turtles.

Based on the linear model, we rejects the null hypothesis with a p value of 2.2e-16. But according to the R^2 value, only 2.857% of the total variance in sex ratio is explained by changes in month. 7 of the 12 months had p-values of less than 0.05, with April, June, August, September, and December being the exception. The p-values of the statistically significant month ranged from < 2e-16 (January) to 0.01576 (February). The impacts of month may be influences by spatial location and local climate. In order to explore monthly sex ratio distribution more clearly, we use box plot to show sex ratio by month for north hemisphere and south hemisphere respectively. The box plot for the north hemisphere clearly shows a trend of more male during colder season (November to April) and female during hotter season (May to October). This result is aligned with conclusion from Question 1 that higher temperature will cause more female turtles. The monthly distribution trend of sex ratio is not clearly for south hemisphere. This maybe because of lack of data in south hemisphere. But the highest proportion of male still appear in the colder season (July and August), and the rest of months all have less than 50% of male. Generally, the month impacts on sex ratio of turtles may also be influenced by the existence of breeding season.

Below is results of linear model and box plots of sex ratio by month for different hemispheres.

```
##
## Call:
## lm(formula = Proportion_Male ~ Month, data = Cleaned_Data_omit)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.6213 -0.4034 -0.0399  0.3882  0.6985
##
```

```
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.47796    0.02046  23.363  < 2e-16 ***
## MonthFeb    -0.11185    0.04631  -2.415  0.01576 *
## MonthMar    -0.17643    0.06084  -2.900  0.00375 **
## MonthApr     0.02323    0.03586   0.648  0.51708
## MonthMay    -0.08198    0.03143  -2.608  0.00913 **
## MonthJun    -0.02139    0.02275  -0.940  0.34708
## MonthJul    -0.14990    0.02975  -5.039 4.86e-07 ***
## MonthAug    -0.00348    0.02879  -0.121  0.90380
## MonthSep    -0.07455    0.03845  -1.939  0.05256 .
## MonthOct    -0.16950    0.05546  -3.056  0.00225 **
## MonthNov     0.14339    0.02833   5.062 4.32e-07 ***
## MonthDec    -0.02247    0.03395  -0.662  0.50803
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4092 on 4366 degrees of freedom
## Multiple R-squared:  0.03101,    Adjusted R-squared:  0.02857
## F-statistic:  12.7 on 11 and 4366 DF,  p-value: < 2.2e-16
```

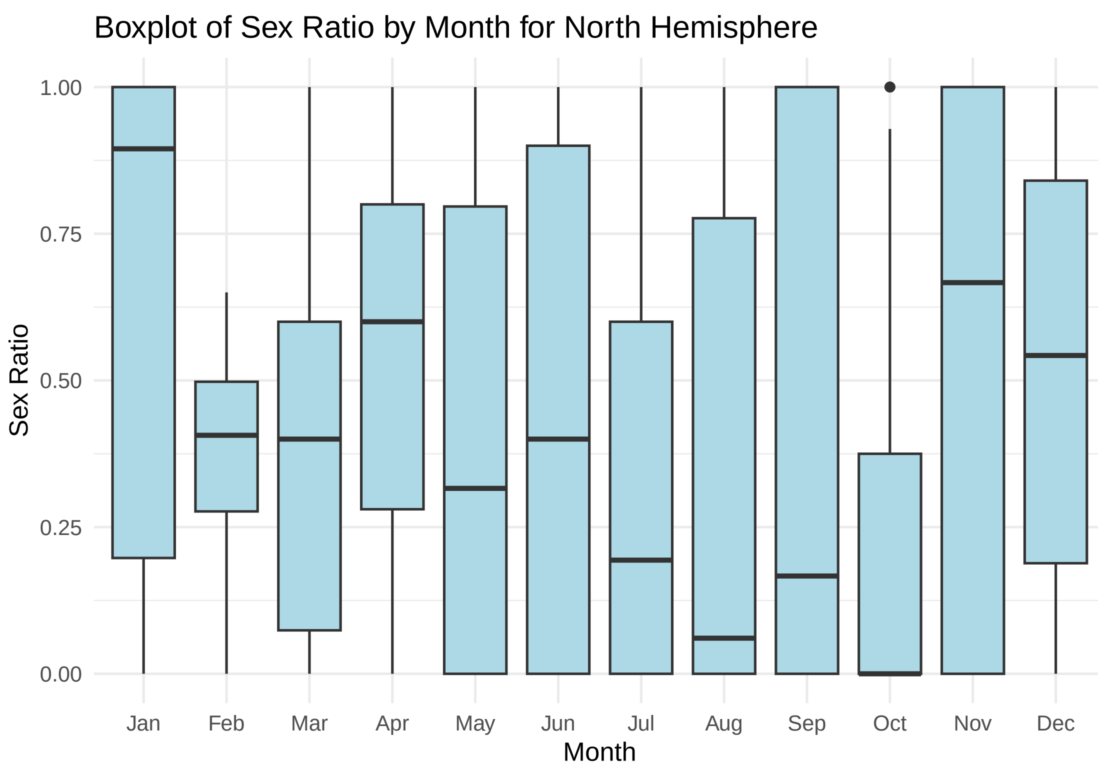

Figure 13: Sex ratio value distribution by month for north hemisphere

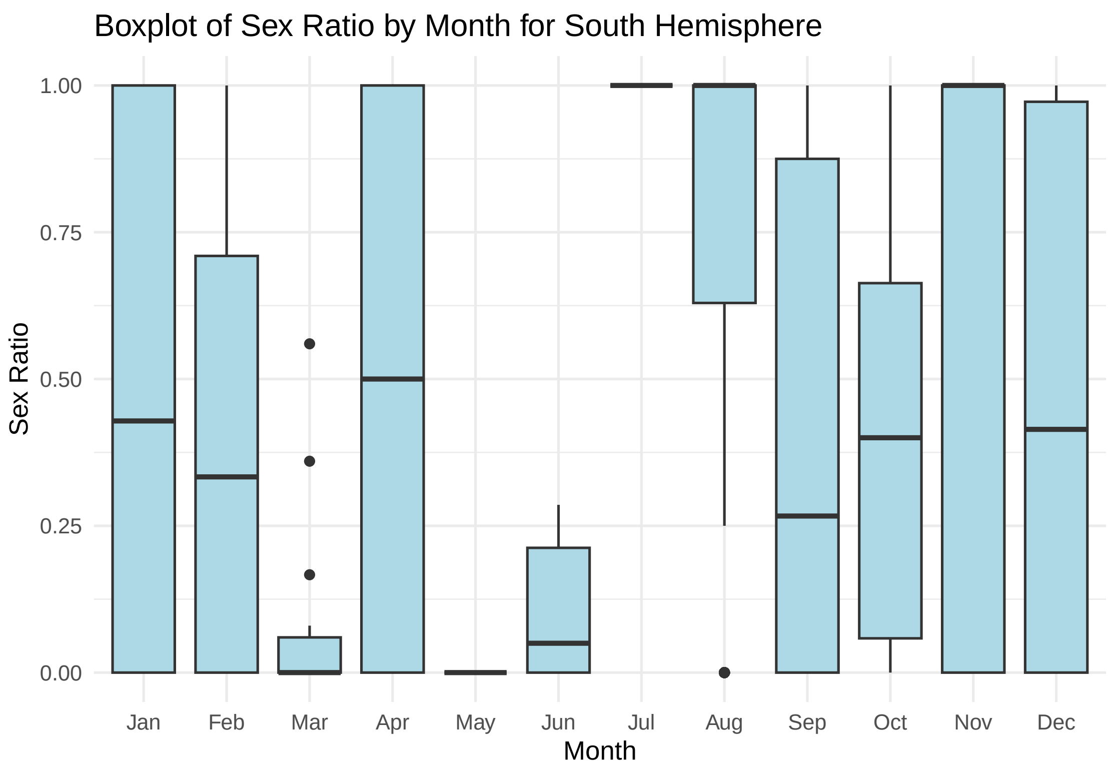

Figure 14: Sex ratio value distribution by month for south hemisphere

4.2.2 Spatial impacts

We created a map to visualize the spatial impact on the sex ratio of turtles. However, it was difficult to identify any distinctive spatial pattern of sex ratio distribution from the map alone. However, after combining the temperature data with the sex ratio distribution, we were able to observe that the distribution of sex ratio is influenced by the temperature at different positions. The results from the map are consistent with the findings from Question 1 that location with lower temperature will have higher male proportion.

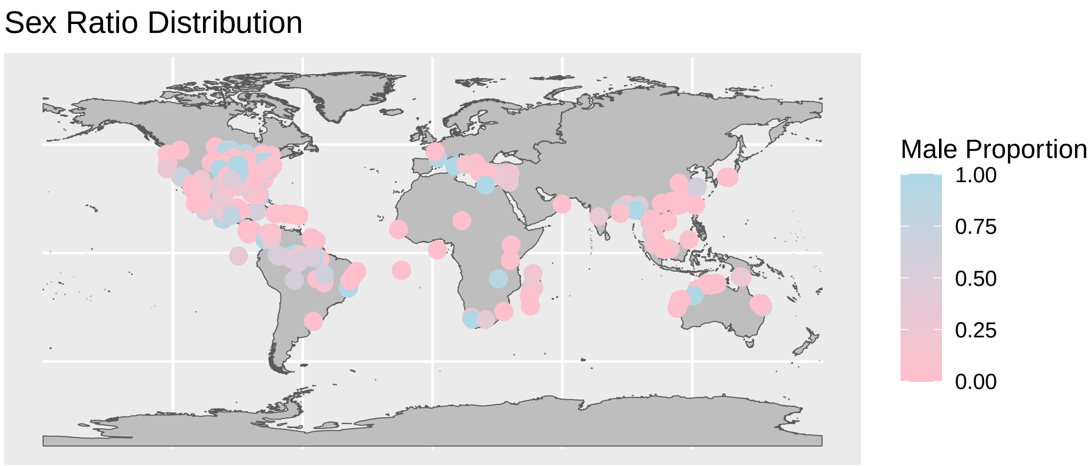

Figure 15: Sex ratio distribution on map

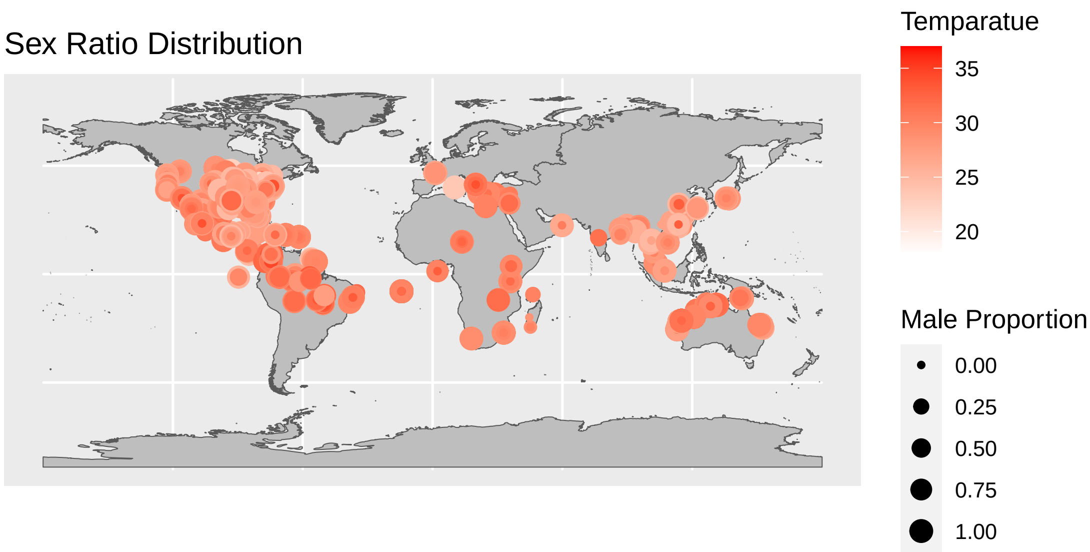

Figure 16: Sex ratio distribution with temperature on map

4.2.3 Best set of explanatory variables

Though month and year do have effects on the sex ratio of turtles, the trend between time and sex ratio may be confounded by the trend between time and temperature. We then want to explore what are best sets of variables for sex ratio determinants. Thus, we run an AIC analysis to determine what set of explanatory variables is best suited to influence sex ratio. We use date data to provide the most detailed temporal information, including day, month, and year, which can capture fine-grained temporal patterns. The result form AIC shows that temperature alone is the best explanatory variables for the sex ratio of turtles. This result may be able to improve by having more valid data as there is only 63 observations due to limited amount of humidity data. Below is the detail of model results.

```
## Start:  AIC=-131.4
## Proportion_Male ~ Mean_Temp + End_Date + Humidity + Captive
##
##             Df Sum of Sq    RSS     AIC
## - Humidity   1   0.00145 1.9449 -133.37
## - End_Date   1   0.01129 1.9548 -133.14
## - Captive    1   0.06095 2.0044 -132.01
## <none>                   1.9435 -131.40
## - Mean_Temp  1   2.42999 4.3735  -96.90
##
## Step:  AIC=-133.36
## Proportion_Male ~ Mean_Temp + End_Date + Captive
##
##             Df Sum of Sq    RSS      AIC
## - End_Date   1   0.08152 2.0265 -133.517
## <none>                   1.9449 -133.365
## - Captive    1   0.09458 2.0395 -133.228
## - Mean_Temp  1   2.48308 4.4280  -98.342
##
## Step:  AIC=-133.52
## Proportion_Male ~ Mean_Temp + Captive
##
##             Df Sum of Sq    RSS     AIC
## - Captive    1   0.05335 2.0798 -134.35
## <none>                   2.0265 -133.52
## - Mean_Temp  1   2.58662 4.6131  -98.50
##
## Step:  AIC=-134.35
## Proportion_Male ~ Mean_Temp
##
##             Df Sum of Sq    RSS      AIC
## <none>                   2.0798 -134.347
## - Mean_Temp  1    2.7131 4.7929  -98.778
```

```
##
## Call:
## lm(formula = Proportion_Male ~ Mean_Temp, data = Cleaned_Data_remove)
##
## Coefficients:
## (Intercept)    Mean_Temp
##      3.4723      -0.1018
```

5 Summary and Conclusions

The sex ratio of turtles can be influenced by temporal factors, although the impact is generally small. While the distribution of sex ratios across space does not exhibit any specific pattern, there is a noticeable trend where locations with lower temperatures tend to have a higher proportion of male turtles. Despite this, it appears that neither spatial nor temporal factors are significant determinants of the sex ratio of turtles. The analysis strongly suggests that temperature is the primary determinant.

The main limitation of our study is that linear regression model may not be the best fit to the data as the trend is more or less polynomial which results in low R-squared values for the linear regression models. Our results may also be limited from the variation of sex ratio among different families and the variation of the amount of data we have for each family. Our data are mainly from north hemisphere which may also affect our results.

Nevertheless, we have seen a strong relationship between temperature and sex ratio of turtles. As climate change gets worse, the increased temperature can easily harm turtles' population by leading to irregular and potentially lethal incubation conditions (NOAA, 2023). Research on how to mitigate the impact of climate change on turtles population and conservation action should be done in the very near future.

Figure 17: Smiling red-eared turtles

6 References

Bull, J. J., & Vogt, R. C. (1979). Temperature-dependent sex determination in turtles. Science, 1186-1188.

Krueger, C. J., & Janzen, F. J. (2022). ROSIE, a database of reptilian offspring sex ratios and sex-determining mechanisms, beginning with Testudines. Scientific Data, 9(1), 22.

NOAA, What causes a sea turtle to be born male or female? (2023). NOAA's National Ocean Service. https://oceanservice.noaa.gov/facts/temperature-dependent.html

Yntema, C. L. (1976). Effects of incubation temperatures on sexual differentiation in the turtle, Chelydra serpentina. Journal of Morphology, 150(2), 453-461.

Photo credit: https://scitechdaily.com/why-does-temperature-determine-the-sex-of-turtles/

https://www.pinterest.jp/pin/61783826112891810/