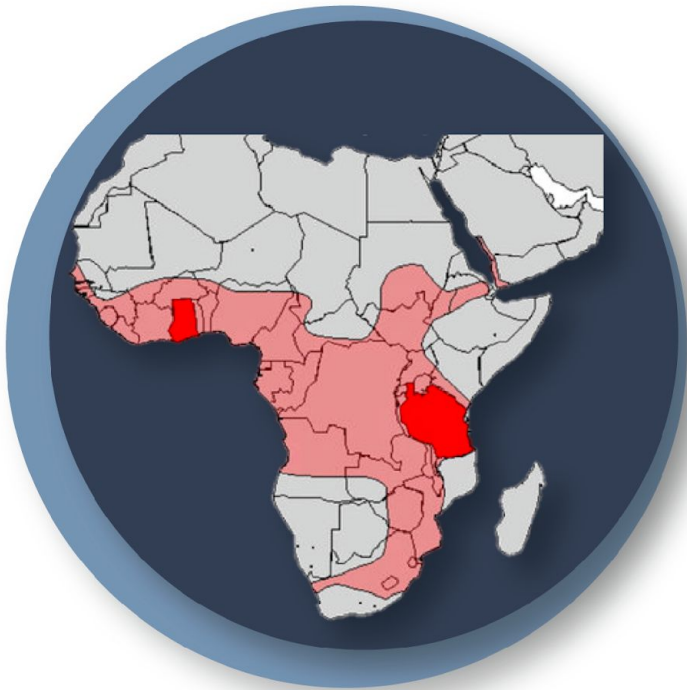# Examining Predictors For Antibodies to Henipavirus-Like Virus in Bats

Ash Malmlov & Katrina Wheeler

EIDOLON HELVUM
AFRICAN STRAW-COLORED FRUIT BAT

- Purported reservoir host for Ebola virus

- Food source in numerous parts of Africa

- Data set:
  Peel, A., Baker, K., Hayman, D. et al. Bat trait, genetic and pathogen data from large-scale investigations of African fruit bats, Eidolon helvum. Sci Data 3, 160049 (2016). https://doi.org/10.1038/sdata.2016.49

| Column header name	Description
Sample	Unique sample identifier, generally consisting of two-letter country identifier, and sample numbe[...]
SamplingEventID	Sampling event identifier, consisting of two-letter country identifier, three-letter location identif[...] format yymmdd
Samplers	Initials of researchers who underto[...] (Andrew C Breed), AFL (Andrés Fe[...] T S Hayman), GCG (Guy-C[...] IVK (Ivan Kuzmin), KSB (K[...] Juste), DJL (DJ Long)[...]
Samplingdate	Sampling date y[...]
Birthdate	Estimated da[...]
Mths.since.birthing	Months (t[...]us birth pulse
Cont_island	Whethe[...]
Country	Count[...] To fa[...]nd Príncipe are part of the country 'São Tomé and Prín[...]
Region	Sam[...]
Location	Spe[...]
Latitude	Latit[...]
Longitude	Long[...]
Bat.wt	Bat we[...]
Sex	Bat sex.[...]
Age	Age classifi[...]J (Juvenile; 2 – <6 months), SI (Sexually Immature; 6 – <24 months) or A[...]
Age	Age classificatio[...]pling relative to the birth pulse. For a subset of samples, the timing of sampling in relati[...]of SI individuals into 6-month age groups SI.1, SI.2 and SI.3 (6 – <12, 12 – <18, 18 – <24 mon[...]
LB[...]VN	Result from Lagos Bat Virus m[...]egative), 1 (seropositive), NA (no result)
Henipavirus	Result from henipavirus Luminex binding assay, using a cutoff so that samples above the cutoff are 99% likely to be seropositive. Data is from the Nipah virus assay, but is presumed to represent cross-reactivity to African henipaviruses rather than the presence of Nipah virus itself. 0 (seronegative), 1 (seropositive), NA (no result)
AchPV1	Result from Achimota virus 1 virus neutralisation assay. 0 (seronegative), 1 (seropositive), NA (no result)
AchPV2	Result from Achimota virus 1 virus neutralisation assay. 0 (seronegative), 1 (seropositive), NA (no result)
GeneticsID	Alternative sample ID used in genetic analyses
Cytb	Cytochrome *b* sequence
GenBankAccession	Accession number for cytochrome *b* sequences deposited in genbank (PopSets 486[...]875 and 73[...]1542)

THE DATA

- Sampled 2827 bats across Africa over 8 years
- Numerous biological samples and measurements
- Focus: Henipavirus

HENIPAH VIRUSES

- Genus comprised of ~20 viruses

- Zoonotic viruses:
  - Hendra virus
    - ~ Australia
    - ~ 70% fatality
  - Nipah virus
    - ~ Indonesia/India
    - ~ 40 to 70% fatality

- Assay used in this dataset
 ◻ Nipah virus
  - Assumed cross-reactivity
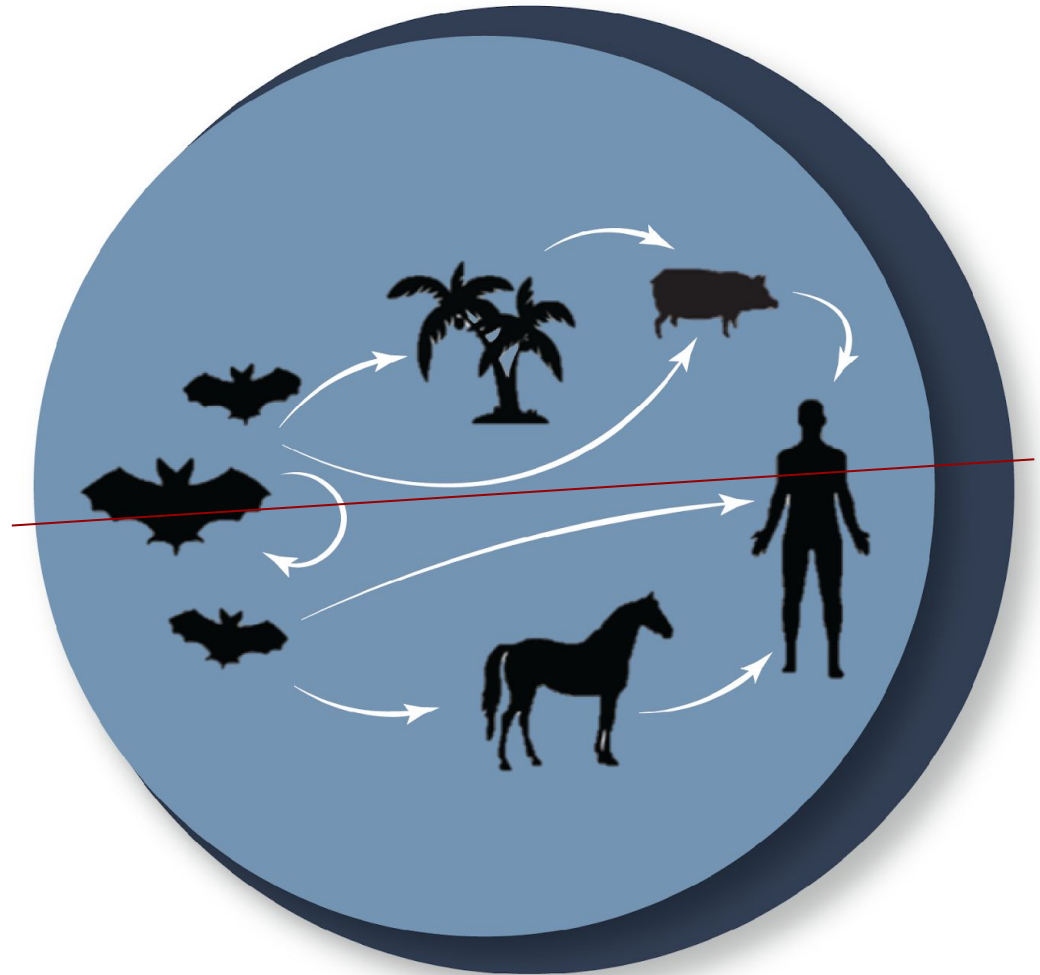
SEROSURVEYS

- Process: collect blood and assay for antibodies to pathogen of interest
    - Specific

- Not assaying for pathogen itself
    - Detecting the immune response to the pathogen
    - Indicates exposure

- Serosurveys are conducted because:
    - Straight-forward
    - Cost effective
    - Looking for specific viruses like looking for a needle in a haystack

- Potential for cross-reactivity with other viruses that share attributes with virus of interest
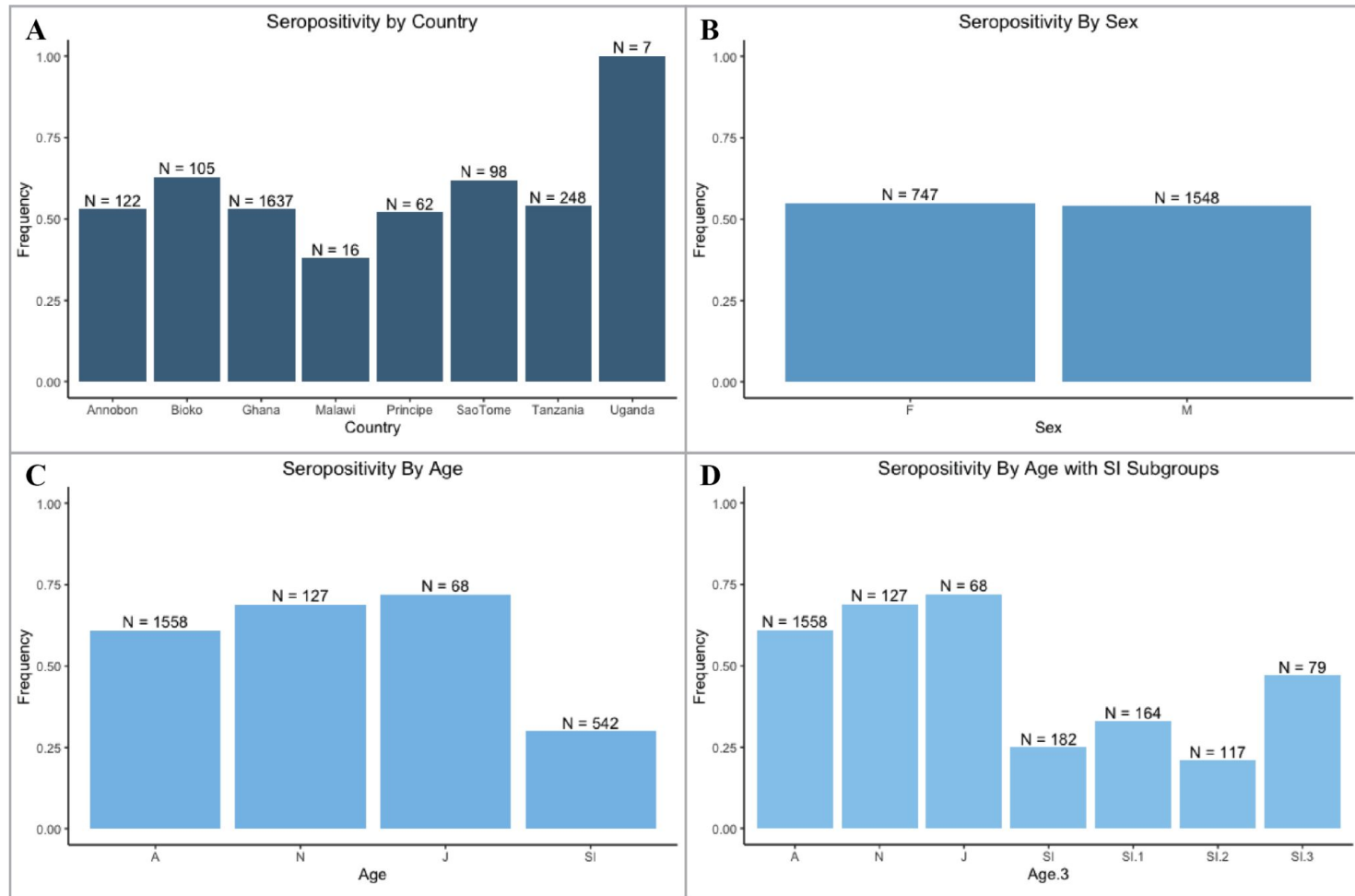
CLEANING DATA

- Dataset contains numerous variables
  - Not all biologically relevant

- Filtered data on those variables that are biologically relevant
  - Sampling date
  - Age and Age3
  - Country
  - Sex

- Omitted NA values

VISUALIZING THE DATA

Number of Seropositive Bats by Date
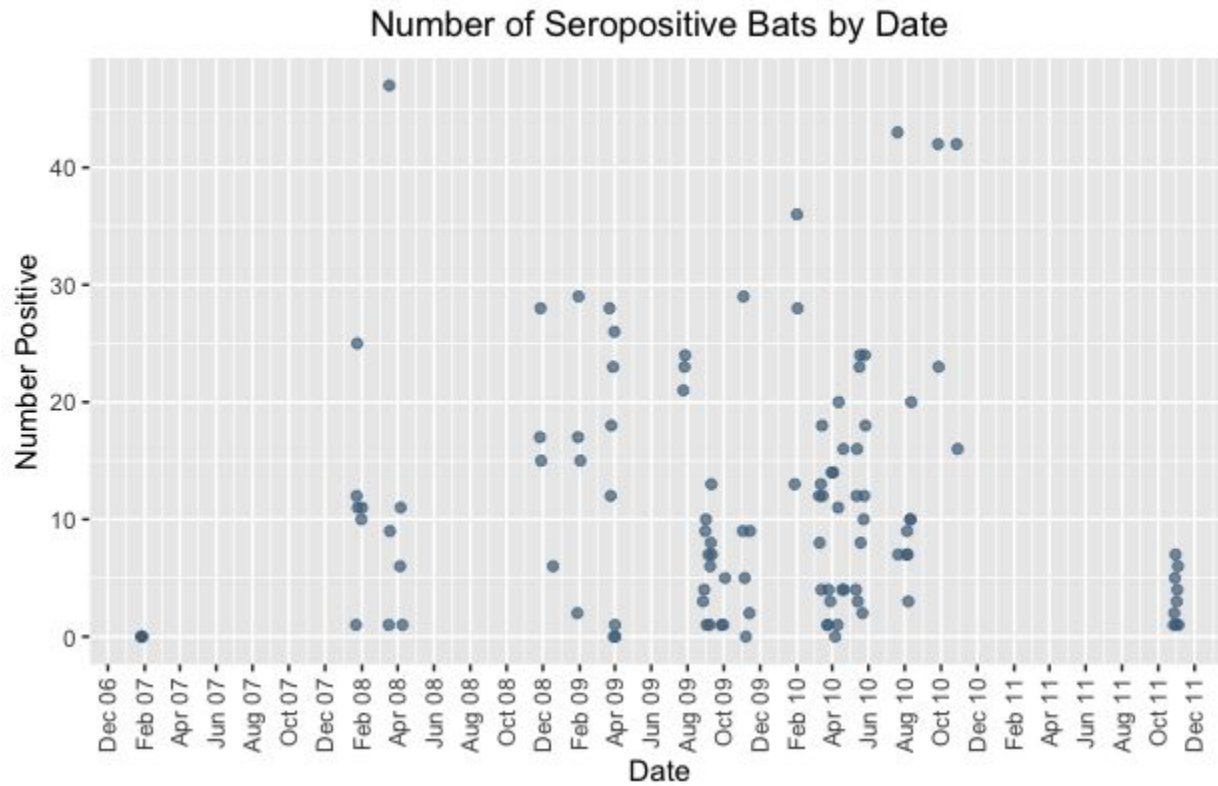
VISUALIZING THE DATA

GENERALIZED LINEAR MODELS

- Broad class of models with 3 components:
    1) Random component (Y): the response variable with which the probability distribution is associated with
    2) Systematic component (X): the explanatory variables as a combination of linear predictors ($X_1, X_2, ..., X_n$).
    3) Link function ($g(u)$): describes the relationship between Y and X

- Logistic Regression:
    - GLM in which Y is binary
    - X can be discrete or continuous
    - $g(u) = \ln\left(\frac{\pi}{1-\pi}\right)$
        - ➢ Log odds

```
logisticAGE <- glm(Henipavirus ~ Age, data = dat.clean, family = "binomial")
summary(logisticAGE)
```

```
Call:
glm(formula = Henipavirus ~ Age, family = "binomial", data = dat.clean)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.5969  -1.3671   0.8698   0.9989   1.5621

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  0.43551    0.05188   8.395   <2e-16 ***
AgeJ         0.51187    0.27519   1.860   0.0629 .
AgeN         0.34152    0.19795   1.725   0.0845 .
AgeSI       -1.30576    0.10751 -12.145   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 3166.0  on 2294  degrees of freedom
Residual deviance: 2984.2  on 2291  degrees of freedom
AIC: 2992.2

Number of Fisher Scoring iterations: 4
```

COEFFICIENTS

```
Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  0.43551    0.05188   8.395   <2e-16 ***
AgeJ         0.51187    0.27519   1.860   0.0629 .
AgeN         0.34152    0.19795   1.725   0.0845 .
AgeSI       -1.30576    0.10751 -12.145   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- Estimates:
    - First row: log odds of the base case
    - Subsequent rows: log odds ratio
    - Base case: determined by alphabetical order

- Standard Error and Z-value:
    - Z-value: estimate/standard error ☐ Wald test

- P – value for Z-value

```
logisticAGE3 <- glm(Henipavirus ~ Age.3, data = dat.clean, family = "binomial")
summary(logisticAGE3)
```

```
Call:
glm(formula = Henipavirus ~ Age.3, family = "binomial", data = dat.clean)

Deviance Residuals:
    Min        1Q    Median        3Q       Max
-1.5969   -1.3671    0.8698    0.9989    1.7800

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)   0.43551    0.05188   8.395  < 2e-16 ***
Age.3J        0.51187    0.27519   1.860   0.0629 .
Age.3N        0.34152    0.19795   1.725   0.0845 .
Age.3SI      -1.54883    0.17948  -8.630  < 2e-16 ***
Age.3SI.1    -1.14701    0.17407  -6.589 4.42e-11 ***
Age.3SI.2    -1.79006    0.23476  -7.625 2.44e-14 ***
Age.3SI.3    -0.56226    0.23136  -2.430   0.0151 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

WANING OF ANTIBODIES IN SI → Age.3

```
Deviance Residuals:
    Min       1Q    Median       3Q       Max
-1.1241  -0.8937  -0.6776   1.2317   1.7800

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)  -0.7115      0.1662  -4.282 1.85e-05 ***
Age.3SI.2    -0.6430      0.2829  -2.273   0.0230 *
Age.3SI.3     0.5847      0.2801   2.088   0.0368 *
---
Signif. codes:  0 '***' 0.001 '**
' ' 1

(Dispersion parameter for binomia

    Null deviance: 451.04  on 359
Residual deviance: 435.78  on 357
AIC: 441.78

Number of Fisher Scoring iterati
```
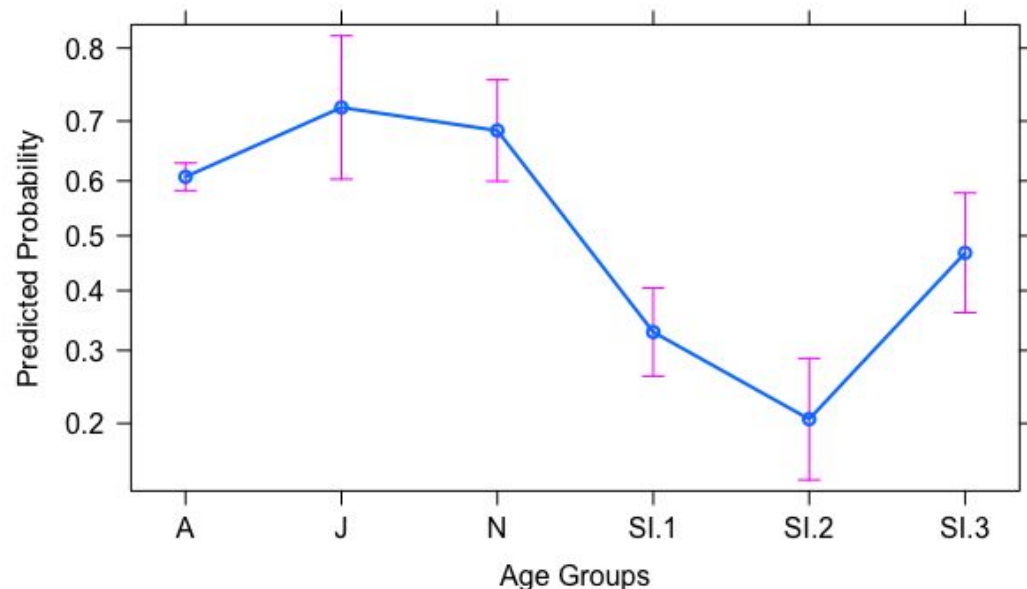

Predicted Probablity of Being Seropositive

BIOLOGICAL INTERPRETATION

- Adults ☐ odds are greater of seroconverstion
  - Broad viral distribution??
  - Confounded by:
    - ~ Antibodies are a long-lived immune response. A seropositive bat may have been exposed 1 month ago or 5 years. Reflection of an accumulation of cases?
    - ~ Potential for cross-reaction may be testing exposure to multiple viruses and not just one

- Neonates/Juveniles☐ passive transfer of maternal antibodies
  - Mammalian trait where offspring acquire antibodies from mom to protect them while immune systems develop
  - Normal for neonates have same antibody repertoire as mom
  - Maternal antibodies wane over time
  - Data indicates maternal antibodies persist through juvenile age group

- Sexually immature ☐ waning of maternal antibodies

ROC Curve & Confusion Matrix

```
Confusion Matrix and Statistics

                Reference
Prediction     seronegative seropositive
  seronegative          382          160
  seropositive          671         1082

               Accuracy : 0.6379
                 95% CI : (0.6179, 0.6576)
    No Information Rate : 0.5412
    P-Value [Acc > NIR] : < 2.2e-16

                  Kappa : 0.2429

 Mcnemar's Test P-Value : < 2.2e-16

            Sensitivity : 0.3628
            Specificity : 0.8712
         Pos Pred Value : 0.7048
         Neg Pred Value : 0.6172
             Prevalence : 0.4588
         Detection Rate : 0.1664
   Detection Prevalence : 0.2362
      Balanced Accuracy : 0.6170

       'Positive' Class : seronegative
```
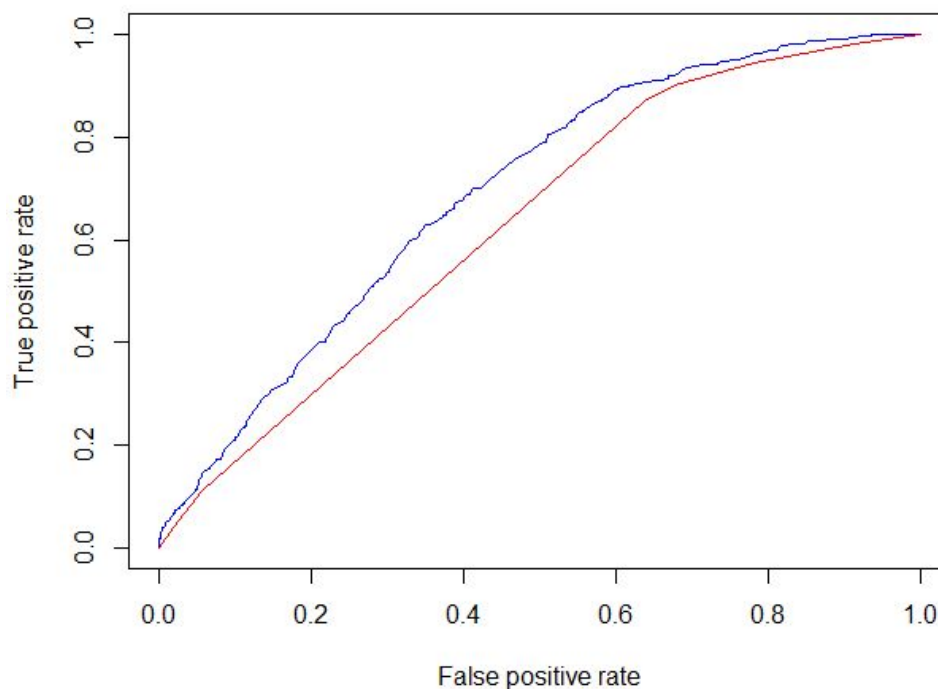
CONCLUSIONS

- The rates of seropositivity were significantly lower in the subcategories of the Sexually Immature age group.
  - o Further research into these subgroups can assist scientists in discovering when young bats are most susceptible to infection.