Chemical Effects on Mammals

https://github.com/c-reents/DataAnalyticsProject

Caroline Reents

Abstract

Data for this study was collected through the EPA ECOTOX database. Data was specifically filtered for effects on mortality, mammals, and chemical subgroup pharmaceutical personal care products. This study aims to determine if different pharmaceutical drugs have different mortality effects for different mammalian species. It also aims to determine if drug class is associated with differing mortality rates. Multiple findings came from these analyses. First species tested was related to the effects on mortality seen, so different organisms reacted to chemicals differently. It is also clear that different drug classes are associated with different mortality outcomes. Finally, I also found that concentration was not associated with different mortality outcomes. These findings have implications in the environmental field, because through the development of new drugs, waterways are becoiming more and more contaminated with pharmaceutical chemicals.

Contents

List of Tables

List of Figures

<Note: set up autoreferencing for figures and tables in your document>

```
## [1] "/Users/carolinereents/Desktop/Data Analytics/EnvironmentalDataAnalytics/FinalPro
```

```
## -- Attaching packages ---------------------------------------------------------
```

```
## v ggplot2 3.1.0     v purrr   0.2.5
## v tibble  1.4.2     v dplyr   0.7.8
## v tidyr   0.8.2     v stringr 1.3.1
## v readr   1.1.1     v forcats 0.3.0
```

```
## -- Conflicts ------------------------------------------------------------------
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
## Loading required package: viridisLite
```

```
## Warning: package 'kableExtra' was built under R version 3.5.2
```

1 Research Question and Rationale

Many pharmaceutical chemicals get put into rivers and streams when they qare excreted from a human body that takes them (americanrivers.org). USGS has found that most streams contain 7 pharmaceutical chemicals in them at all times (americanrivers.org). Many animals are exposed to these sources of water and therefore any chemicals within them. Because of this phenomenon, I thought it would be interesting to see if different types of drugs effected small mammals differently. With these drugss becoming more and more prolific in waters across the nation, this will increasingly become an issue.

This data set includes the effects of certain pharmaceutical or health care drugs or chemicals on the mortality of various mammals. Each data point is related to a specific study done on that chemical with a certain mammal. Mortality is measured in various "endpoints" which refer to how much mortality was seen in each study. There are 24 different chemicals in this data set, large groups of these chemicals are either cancer-combating drugs or antibiotics. For this project I aim to determine if different mammals are more suscepitble to death from being exposed to these chemicals, or if mortality is highly dependent on concentration of chemical. I also want to determine if the cancer drugs and the antibiotics differ in their likelihood to cause death.

2 Dataset Information

The following is a list of all chemicals tested in this data set pulled from the ECOTOX webpage. Each data point is the result of a study, so each data point seen on visualizations represents the findings of one or a group of researchers published study.

The data was pulled filtering for mammals, effects of mortality, all endpoints, and chemical subgroup "pharmaceutical personal care products". The drugs tested in this data set were:

- 17alpha-ethynylestradiol : used in BC
- 6-mercaptopurine : used for cancer and autoimmune
- Ampicillin - antibiotic
- Betamethasone - steroid
- Clindamyacin - antibiotic
- Dexamethasone - steroid
- Dexamethasone sodium - steroid
- Diazepam - pain killer
- Dimenhydrinate - motion sickness
- Erythomycin - antibiotic
- Gentamycin - antibiotic
- Lindane - treats scabies
- Methimazole - treats hyperthyroid
- Methotrexate - cancer autoimmune treatment
- Metronidazole - antibiotic
- Phenobarbital - sedative
- Prednisone - steroid
- Propranolol - beta-blocker
- Tetracycline - antibiotic
- Tetracycline hydrochloride - antibiotic
- Thalidomide - cancer treeatment
- Trans-Retinoic acid - cancer
- Triamcinolone - treats inflammation
- Warfarin - prevents blood clots

It is important to note that the endpoints were defined as follows:

- EC10, EC50: effective concentration to x % of test organisms
- ET50: effective response time to 50 % of test organisms
- LC0 through LC95: lethal concentration to x % of test organisms
- LD50, LD95, LD99: lethal dose to x % of test organisms
- LOEC: lowest observable effect concentration
- LOEL: lowest observable effect level
- LT50: time to 50 % mortality of organisms
- NOEC: no observable effect concentration
- NOEL: no observable effect level
- NR: not reported

- NR-LETH: 100 % mortality or 0 % survival of test organisms
- NR-ZERO: 0 % mortality or 100 % survival of test organisms

	Chemical.Name	Drug.Class	Species.Name	Common.Name
2	6-Mercaptopurine	Anti-cancer	Ochotona rufescens ssp. rufescens	Afghan Pika
3	6-Mercaptopurine	Anti-cancer	Ochotona rufescens ssp. rufescens	Afghan Pika
4	Ampicillin	Antibiotic	Ochotona rufescens ssp. rufescens	Afghan Pika
5	Ampicillin	Antibiotic	Ochotona rufescens ssp. rufescens	Afghan Pika

Note that this table wass to large to fit on the page, so it is currently showing rows 2-5 and columns 2-5

3 Exploratory Data Analysis and Wrangling

```
#summary code
colnames(Mammal_dat_RAW)
```

```
##  [1] "CAS.No."           "Chemical.Name"     "Drug.Class"
##  [4] "Species.Name"      "Common.Name"       "Effect"
##  [7] "Measurement"       "Endpoint"          "Dur..Std."
## [10] "Conc..Type"        "Conc..Mean..Std."  "Conc..Units..Std."
## [13] "Pub..Year"         "Citation"
```

```
class(Mammal_dat_RAW$Pub..Year)
```

```
## [1] "integer"
```

```
summary(Mammal_dat_RAW$Common.Name)
```

```
##              Afghan Pika             American Mink                Black Rat
##                       15                         4                        9
## Black-Tailed Prairie Dog               Deer Mouse           Domestic Sheep
##                        1                         2                        1
##           Golden Hamster               House Mouse              Meadow Vole
##                        5                        37                        1
##               Norway Rat           Polynesian Rat           Tammar Wallaby
##                       18                         1                        3
```

```
summary(Mammal_dat_RAW$Chemical.Name)
```

```
##       17alpha-Ethinylestradiol                6-Mercaptopurine
##                              1                               2
##                     Ampicillin                  Betamethasone
##                              3                               2
##                    Clindamycin                  Dexamethasone
##                              1                               5
## Dexamethasone sodium phosphate                       Diazepam
##                              3                               1
##                  Dimenhydrinate                   Erythromycin
##                              1                               1
##                     Gentamycin                        Lindane
##                              1                               1
##                    Methimazole                   Methotrexate
##                              1                               1
##                  Metronidazole                  Phenobarbital
##                              1                               4
##                     Prednisone                    Propranolol
##                              2                               1
```

```
##                   Tetracycline     Tetracycline hydrochloride
##                              7                             28
##                    Thalidomide           trans-Retinoic acid
##                             10                              2
##                   Triamcinolone                       Warfarin
##                              2                             16
```

```
class(Mammal_dat_RAW$Chemical.Name)
```

```
## [1] "factor"
```

```
#Thalidomide and warfarin have most data points

class(Mammal_dat_RAW$Conc..Mean..Std.)
```

```
## [1] "numeric"
```

```
Mammal_dat_RAW$Conc..Mean..Std.<- as.integer(Mammal_dat_RAW$Conc..Mean..Std.)
#changed cancentration to an integer rather than numeric

summary(Mammal_dat_RAW$Pub..Year)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    1950    1976    1986    1984    1989    2014
```

```
Mammal_dat_RAW$Pub..Year<-as.factor(Mammal_dat_RAW$Pub..Year)
summary(Mammal_dat_RAW$Pub..Year)
```

```
## 1950 1964 1965 1966 1967 1976 1977 1978 1982 1984 1985 1986 1987 1989 1990
##    1    4    7    1    1   12    2    5    1    1    2   15   13    8    2
## 1991 1992 1995 1996 1999 2002 2006 2013 2014
##    1    8    1    3    4    1    1    1    2
```

```
#treated this as a factor so that I could see how many studies took place in each year

summary(Mammal_dat_RAW$Endpoint)
```

```
##    LC50    LD50    LOEL    NOEL      NR NR-LETH NR-ZERO
##       6      20       2      21       5       7      36
```

```
class(Mammal_dat_RAW$Endpoint)
```

```
## [1] "factor"
```

```
ECOTOX_LethZero <- Mammal_dat_RAW %>%
  filter(Endpoint == "NR-LETH" | Endpoint =="NR-ZERO")

ECOTOX_Lethal <- Mammal_dat_RAW %>%
  filter(Endpoint == "NR-LETH")
```

```r
class(ECOTOX_LethZero$Endpoint)
```

```
## [1] "factor"
```

```r
ECOTOX_LethZero$Endpoint<-as.character(ECOTOX_LethZero$Endpoint)

summary(Mammal_dat_RAW$Chemical.Name)
```

```
##       17alpha-Ethinylestradiol                6-Mercaptopurine
##                              1                               2
##                     Ampicillin                   Betamethasone
##                              3                               2
##                    Clindamycin                   Dexamethasone
##                              1                               5
## Dexamethasone sodium phosphate                        Diazepam
##                              3                               1
##                  Dimenhydrinate                     Erythromycin
##                              1                               1
##                     Gentamycin                         Lindane
##                              1                               1
##                     Methimazole                     Methotrexate
##                              1                               1
##                   Metronidazole                   Phenobarbital
##                              1                               4
##                     Prednisone                     Propranolol
##                              2                               1
##                    Tetracycline     Tetracycline hydrochloride
##                              7                              28
##                    Thalidomide            trans-Retinoic acid
##                             10                               2
##                   Triamcinolone                        Warfarin
##                              2                              16
```

```r
ECOTOX_antibiotics<- Mammal_dat_RAW %>%
  filter(Chemical.Name == "Ampicillin"|
           Chemical.Name == "Clindamycin"|
           Chemical.Name =="Gentamycin"|
           Chemical.Name ==  "Metronidazole"|
           Chemical.Name == "Erythromycin"|
           Chemical.Name == "Tetracycline"|
           Chemical.Name == "Tetracycline hydrochloride")
View(ECOTOX_antibiotics)
```

```r
#wrangle the data to only have the three most commonly studied organisms

Three_dat <- Mammal_dat_RAW %>%
  filter(Common.Name == "Afghan Pika" | Common.Name == "Norway Rat" | Common.Name =="Hou

summary(Mammal_dat_RAW$Chemical.Name)
```

```
##       17alpha-Ethinylestradiol                6-Mercaptopurine
##                              1                               2
##                    Ampicillin                   Betamethasone
##                              3                               2
##                    Clindamycin                   Dexamethasone
##                              1                               5
## Dexamethasone sodium phosphate                        Diazepam
##                              3                               1
##                  Dimenhydrinate                    Erythromycin
##                              1                               1
##                     Gentamycin                         Lindane
##                              1                               1
##                     Methimazole                    Methotrexate
##                              1                               1
##                   Metronidazole                   Phenobarbital
##                              1                               4
##                     Prednisone                     Propranolol
##                              2                               1
##                    Tetracycline     Tetracycline hydrochloride
##                              7                              28
##                    Thalidomide            trans-Retinoic acid
##                             10                               2
##                   Triamcinolone                        Warfarin
##                              2                              16
```

```r
#wrangle data to combine endpoints categories = 0, 50, 100


Three_dat_mutateMortality<-mutate(Three_dat,
              Mortality.Rate =
              ifelse(Three_dat$Endpoint=="LD50", "Lethal", "Not Lethal"),
              ifelse(Three_dat$Endpoint=="LOEL", "Not Lethal", "Lethal"),
              ifelse(Three_dat$Endpoint=="NOEL", "Not Lethal", "Lethal"),
              ifelse(Three_dat$Endpoint=="NR-LETH", "Lethal", "Not Lethal"),
              ifelse(Three_dat$Endpoint=="NR-ZERO", "Not Lethal", "Lethal"))
```

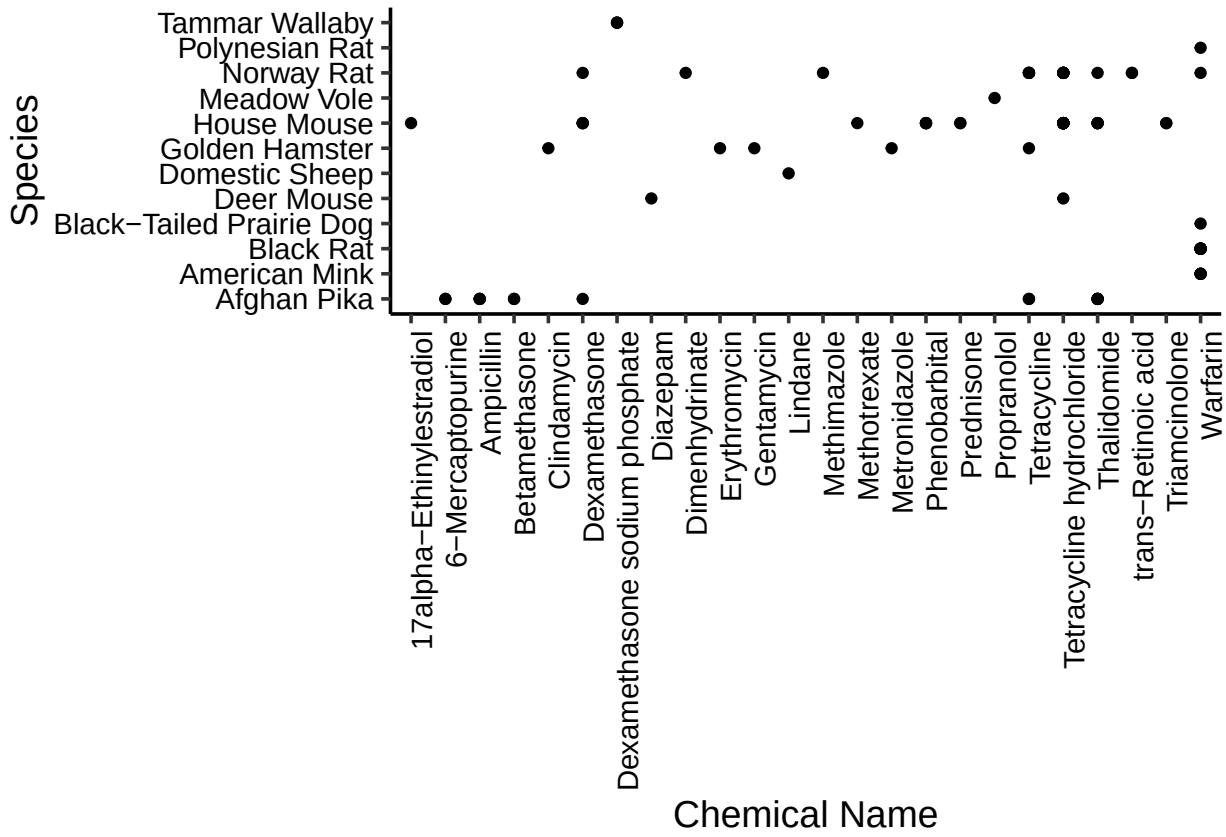

Figure 1: Figure 1: Exploratory visualization 1, species tested for each chemcial

```
ECOTOX_cancerdrugs<- Mammal_dat_RAW %>%
  filter(Chemical.Name == "6-Mercaptopurine"|
           Chemical.Name == "Methotrexate"|
           Chemical.Name =="Thalidomide"|
           Chemical.Name =="trans-Retinoic acid")
```

The first exploratory graph shows which drugs were tested on which animals. This helps visualize if each drug was tested on each animal, which could effect the results of certain statistical tests. For the second exploratory figure, I wanted to see which drugs resulted in which endpoints. I was particulary interested to see which drugs resulted in 100% mortality (NR-LETH). For the third exploratory figure, I determined the spread of concentration for the drugs that caused 100% mortality (NR-LETH). I wanted to determine if certain drugs required higher concentrations to kill and if certain drugs killed at multiple concentrations.

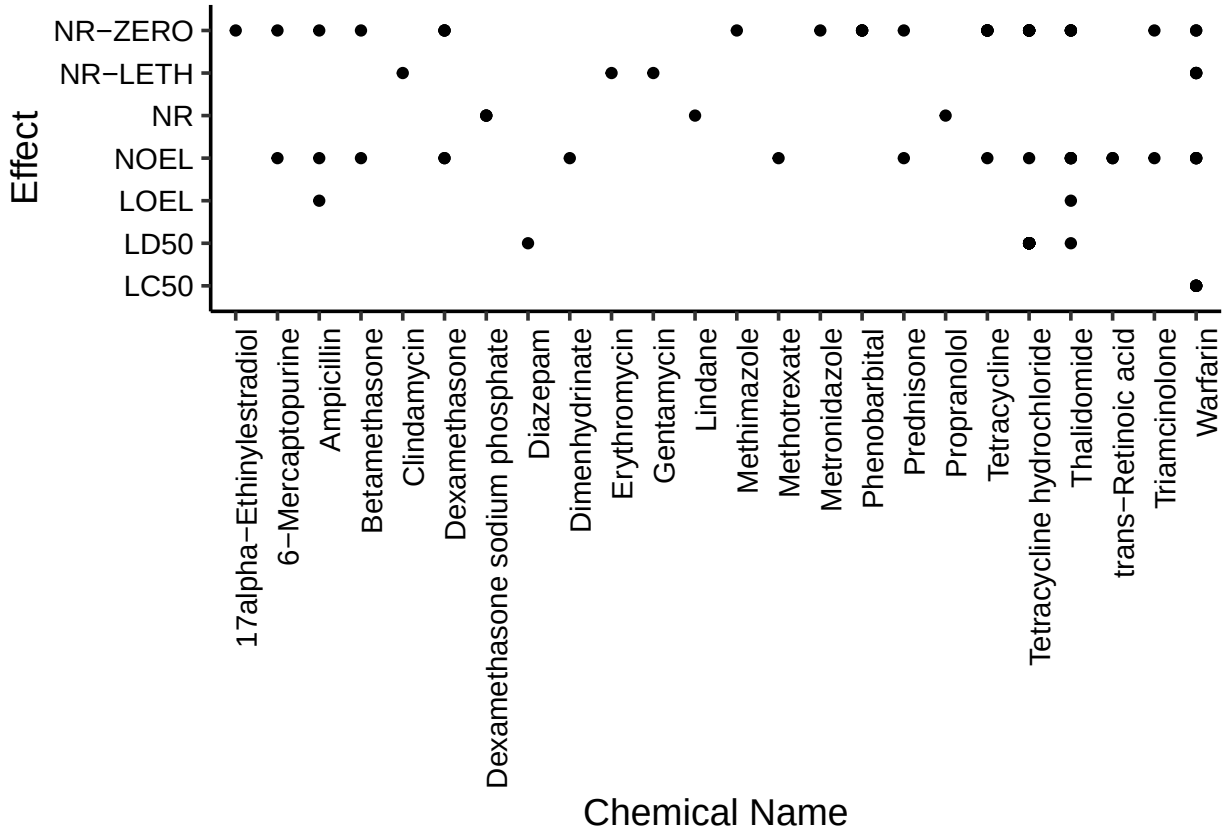

Figure 2: Figure 2: Exploratory visualization 2, endpoint effects for each chemical

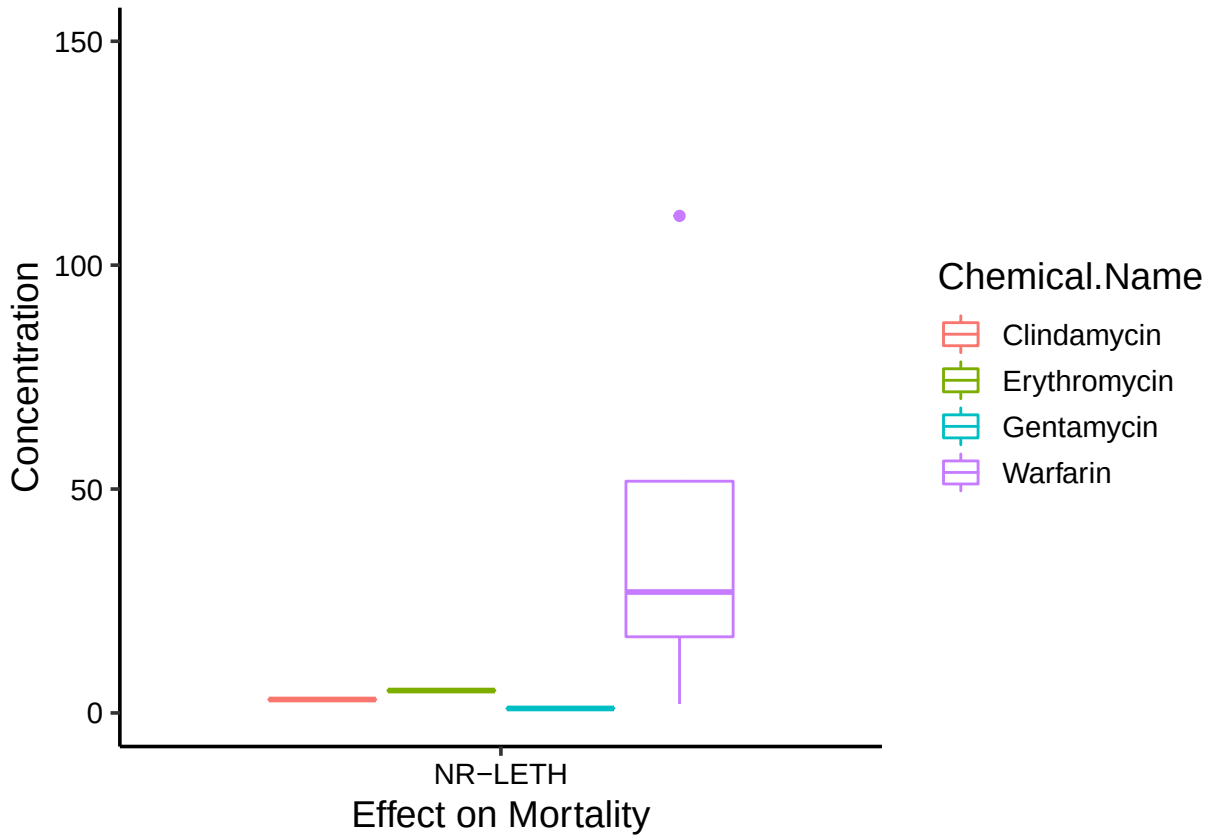

Figure 3: Figure 3: Exploratory visualization 3, concentration of drugs assocaited with 100% mortality

4 Analysis

```
chisq.test(table(Mammal_dat_RAW$Drug.Class, Mammal_dat_RAW$Endpoint), correct = FALSE)
```

```
## Warning in chisq.test(table(Mammal_dat_RAW$Drug.Class,
## Mammal_dat_RAW$Endpoint), : Chi-squared approximation may be incorrect
```

```
##
##  Pearson's Chi-squared test
##
## data:  table(Mammal_dat_RAW$Drug.Class, Mammal_dat_RAW$Endpoint)
## X-squared = 48.352, df = 12, p-value = 2.714e-06
```

```
chisq.test(table(Mammal_dat_RAW$Common.Name, Mammal_dat_RAW$Endpoint), correct=FALSE)
```

```
## Warning in chisq.test(table(Mammal_dat_RAW$Common.Name,
## Mammal_dat_RAW$Endpoint), : Chi-squared approximation may be incorrect
```

```
##
##  Pearson's Chi-squared test
##
## data:  table(Mammal_dat_RAW$Common.Name, Mammal_dat_RAW$Endpoint)
## X-squared = 228.26, df = 66, p-value < 2.2e-16
```

```
mylogit <- glm(Endpoint ~ Conc..Mean..Std., data = Mammal_dat_RAW, family = "binomial")
```

```
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
```

```
summary(mylogit)
```

```
##
## Call:
## glm(formula = Endpoint ~ Conc..Mean..Std., family = "binomial",
##     data = Mammal_dat_RAW)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -2.45219   0.00522   0.07686   0.52885   0.61042
##
## Coefficients:
##                  Estimate Std. Error z value Pr(>|z|)
## (Intercept)       1.58576    0.50945   3.113  0.00185 **
## Conc..Mean..Std.  0.01413    0.01126   1.255  0.20956
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
```

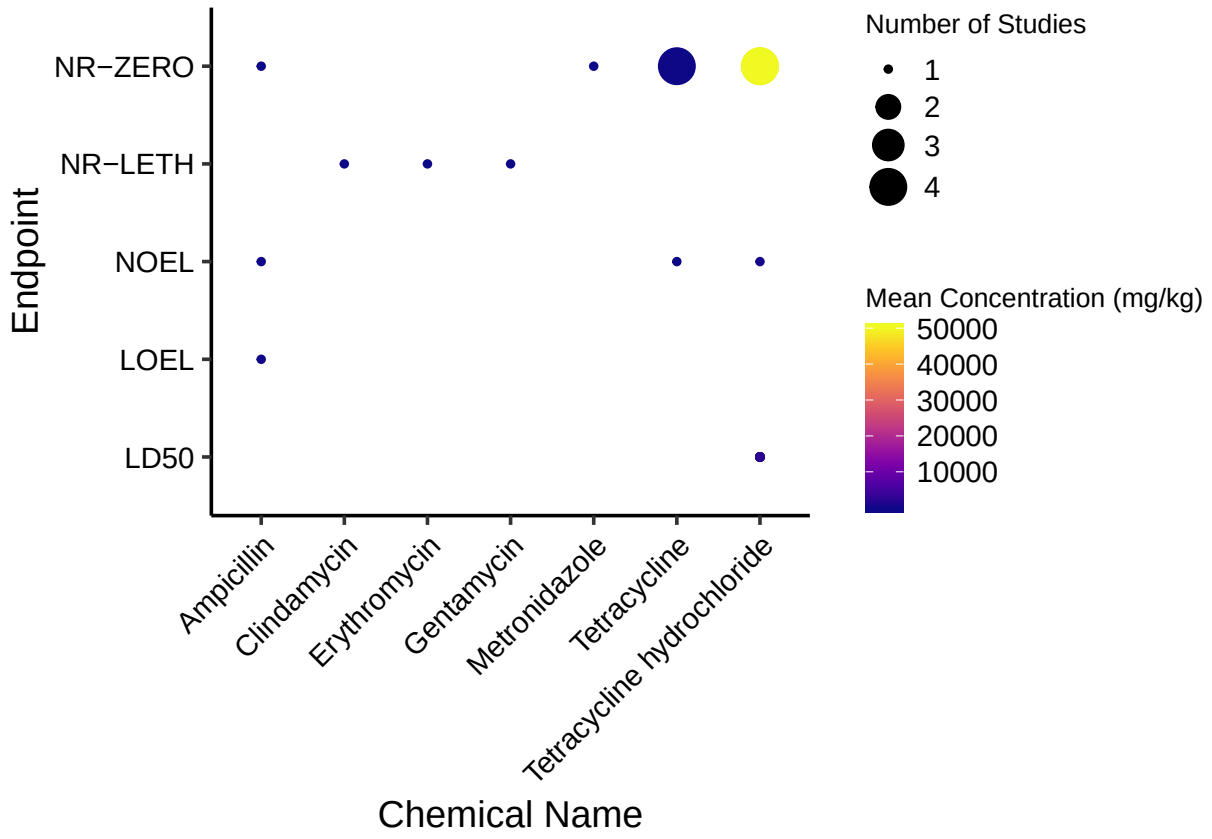

Figure 4: Figure 4: Antibiotics and their endpoints based on concentration. Larger dots indicate more studies.

```
##
##     Null deviance: 45.016  on 96  degrees of freedom
## Residual deviance: 35.578  on 95  degrees of freedom
## AIC: 39.578
##
## Number of Fisher Scoring iterations: 14
```

It was difficult to run statistical tests on this data because I was working with mostly categorical data. This is why I chose to run chi squared tests and a logit glm. While the chi squared test warns that it may not be accurate due to sample size, I still ran it and cited it as the findings were seen in the visual aides. That said, it is important to note that these finidings may nt be statistically significant, which is probably because many drugs or species had limited studies done on them. I had to use a binomial glm because the dependent variable was categorical.

```
## Warning: Ignoring unknown parameters: breaks
```

Figure 4+5: The goal of visualization one and two is to visualize both the concentrations of the different drugs associated with the different endpoints, but also how many studies were done on that drug. Presumably the bigger the dot the more robust that data point because it is based off of many studies instead of just one.

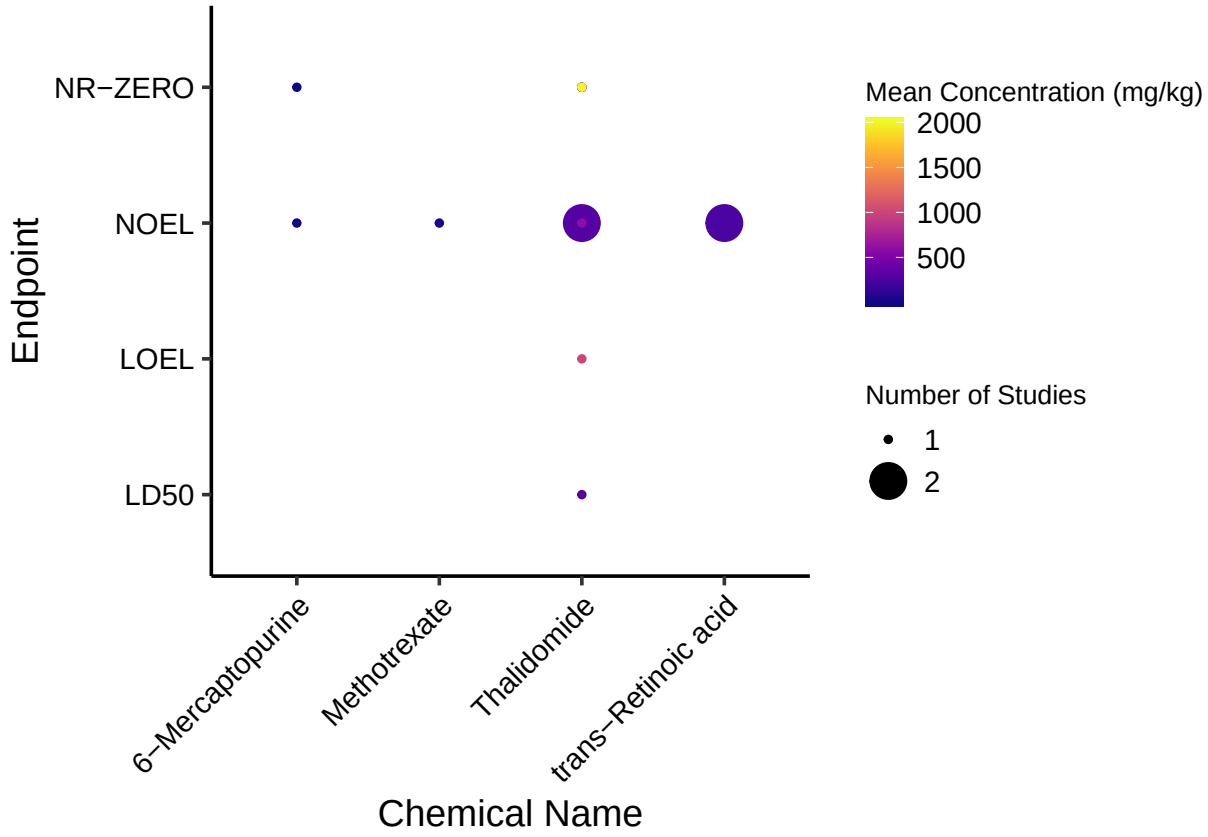

Figure 5: Figure 5: Anti-cancer drugs and their endpoints based on concentration. Larger dots indicate more studies.

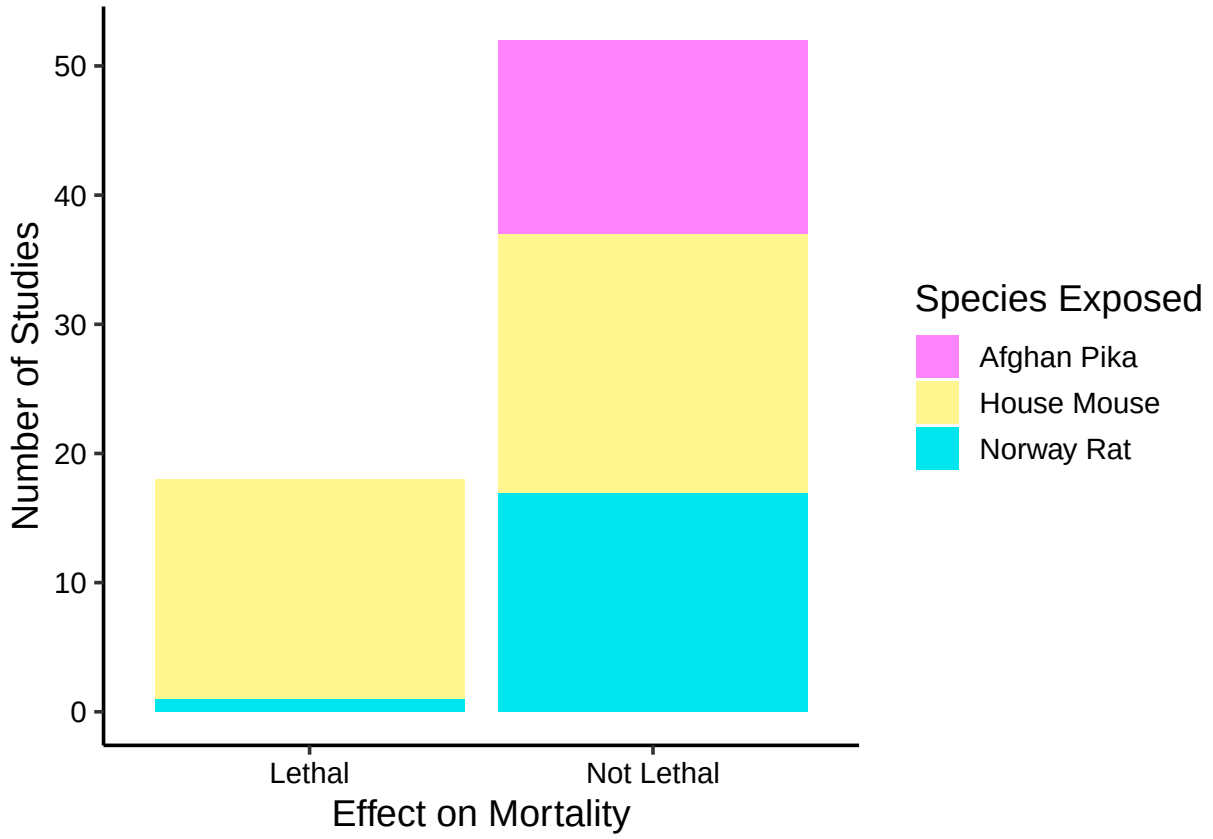

Figure 6: Figure 6: Top three most studied organisms and the number of studies which proved lethal for that organism. 'Lethal' includes LD50 (50% morality) and NR-LETH (100% mortality), 'Not Lethal' includes measures with minimal or no effect on the organism.

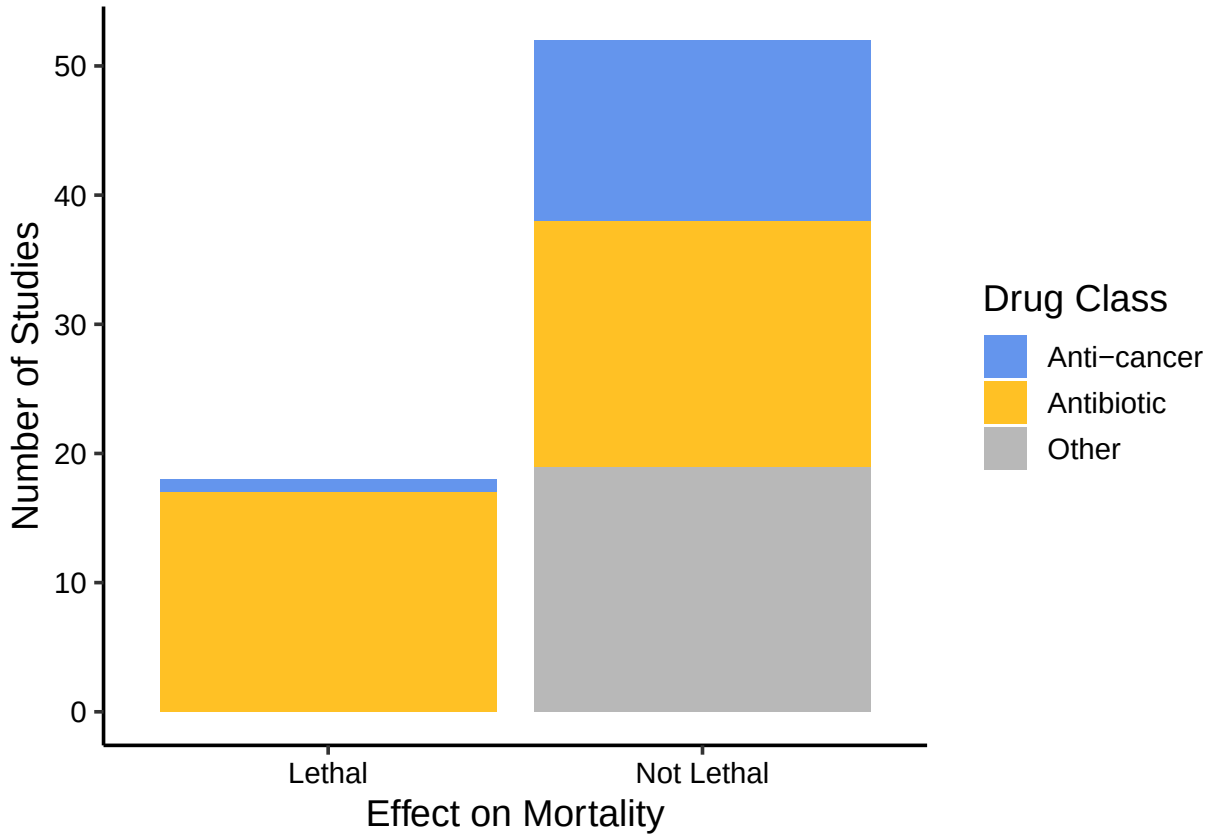

Figure 7: Figure 7: Number of studies associated with each drug class and whether or not they proved lethal. Again, 'Lethal' includes LD50 (50% morality) and NR-LETH (100% mortality), 'Not Lethal' includes measures with minimal or no effect on the organism.

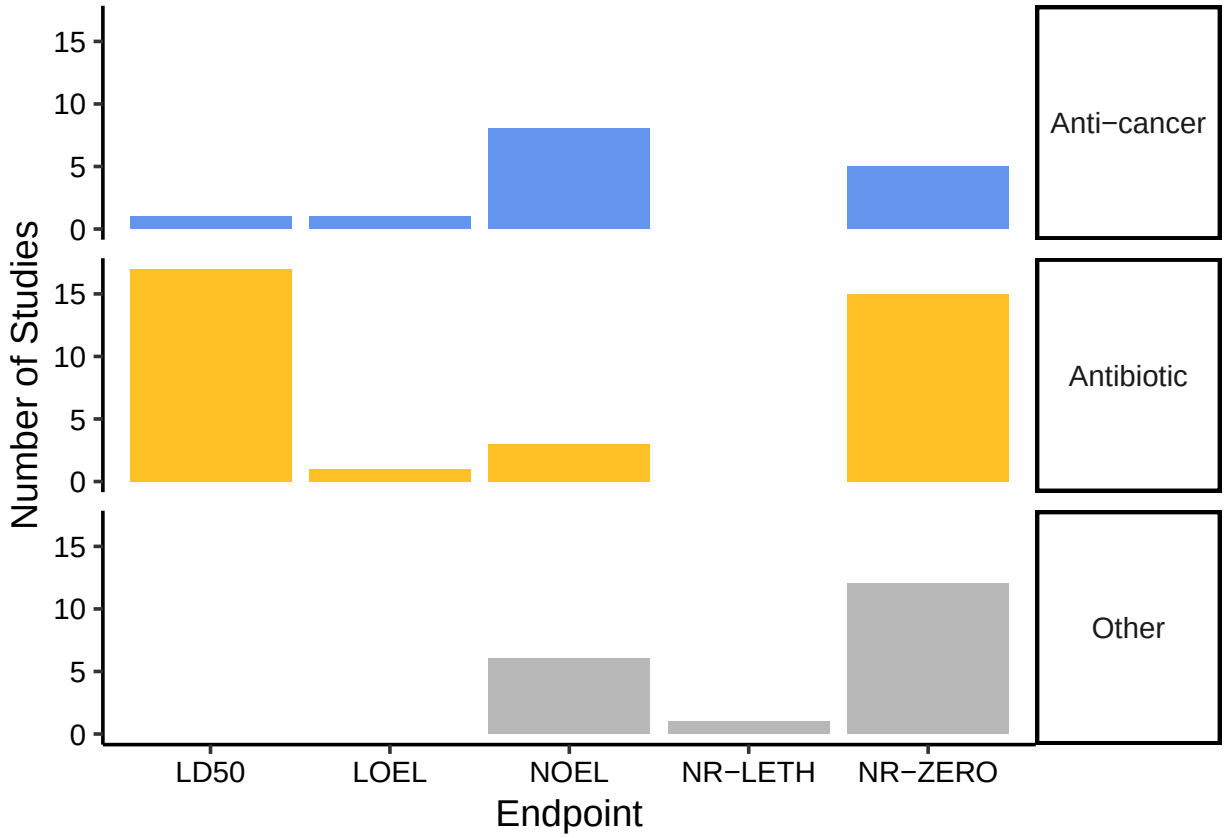

Figure 8: Figure 8: Endpoint effects of each drug class dependent on how many studies came to the same conclusion.

Figure 6: The goal here is to briefly visualize the proportion of studies that were lethal to which of the three most tested organisms. I excluded animals that were not studied as often, so that the results are more robust.

Figure 7:The goal of this graph is similar to the previous one, but I wanted to visualize the proportion of studies found to be lethal or not for each drug class. While it is not based on proportions, it is still a good visual to gague how many studies total found each outcome and which drugs were tested in those studies.

Figure 8: In this figure I wanted to further break down and see each end point seperately. Mainly because there is an important difference from 50% mortality to 100% mortality.

5 Summary and Conclusions

Through my statistical tests, I found a few patterns in the data. By running a chi squared test to determine the relation between drug class and endpoint, I found that the class of drug is related to the endpoint ($p<0.05$). This means that whether a drug is an antibiotic, anti-cancer drug, or "other" is associated with the morotality of the test organsims. By looking at (Figure 7), you can see that of the studies finding lethal effects, most were antibiotics.

Another chi squared test found that the relationship between species tested and mortality rates was also significantly associated ($p<0.05$). This means that certain species experienced different effects than others. Of course, this may be just due to the fact that each drug was not tested on each organism. By looking at (Figure 6), you can see that of the studies that found lethal effects, most were tested on the common house mouse.

Through running a logit model, I did find that concentration was not a significant predictor of endpoint ($p>0.05$). Applying this to the issue of water contamination, this may mean that concentrations will not effect how contamination effects animals. Depending on the drug this maay be good or bad. This may mean that small doses of chemicals have the same effect as large doses, so while we don't need to worry about concentrations getting higher, it is clear that we need to worry about the concentrations that are already present. Looking at Figure 4 you can see that most drugs were test in low concentrations. It is interesting to note that the top two most studied drugs had zero mortality and the ones that proved lethal had less studies associated with them. Loojing at tetracycline hyrdrochloride I find it interesting that a lower concentration had the LD50 effect, which indicates 50% mortality, while the higher concentration was associated with zero mortality. By loking at Figure 5 you can see that none of the anti-cancer drugs caused 100% mortality, the only lethality was seen in one study on Thalidomide (LD50- 50% mortality). Looking at Figure 8 you can see again that antibitics causes the most lethal outcomes. I was surprised by these findings, I expected that anti-cancer drugs, like Methotrexate would have more lethal effects than antibiotics.