Urine Cultures 2019

```
## -- Attaching packages --------------------------------------------------------- tidyverse 1.2.1 --

## v ggplot2 3.1.0     v purrr   0.2.5
## v tibble  2.0.1     v dplyr   0.7.8
## v tidyr   0.8.2     v stringr 1.3.1
## v readr   1.3.1     v forcats 0.3.0

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

##
## Attaching package: 'data.table'

## The following objects are masked from 'package:dplyr':
##
##     between, first, last

## The following object is masked from 'package:purrr':
##
##     transpose

##
## Attaching package: 'lubridate'

## The following objects are masked from 'package:data.table':
##
##     hour, isoweek, mday, minute, month, quarter, second, wday,
##     week, yday, year

## The following object is masked from 'package:base':
##
##     date

##
## Attaching package: 'gridExtra'

## The following object is masked from 'package:dplyr':
##
##     combine

## rgdal: version: 1.3-6, (SVN revision 773)
##  Geospatial Data Abstraction Library extensions to R successfully loaded
##  Loaded GDAL runtime: GDAL 2.1.3, released 2017/20/01
##  Path to GDAL shared files: /Library/Frameworks/R.framework/Versions/3.5/Resources/library/rgdal/gda
##  GDAL binary built with GEOS: FALSE
##  Loaded PROJ.4 runtime: Rel. 4.9.3, 15 August 2016, [PJ_VERSION: 493]
##  Path to PROJ.4 shared files: /Library/Frameworks/R.framework/Versions/3.5/Resources/library/rgdal/p
##  Linking to sp version: 1.3-1
```

```
## Checking rgeos availability: FALSE
##      Note: when rgeos is not available, polygon geometry     computations in maptools depend on gpcli
##      which has a restricted licence. It is disabled by default;
##      to enable gpclib, type gpclibPermit()


##
## Spatial Point Pattern Analysis Code in S-Plus
##
##  Version 2 - Spatial and Space-Time analysis


##
## Attaching package: 'splancs'

## The following object is masked from 'package:dplyr':
##
##      tribble


## The following object is masked from 'package:tibble':
##
##      tribble


## Loading required package: gpclib


## General Polygon Clipper Library for R (version 1.5-5)
##  Type 'class ? gpc.poly' for help


## [1] TRUE
```
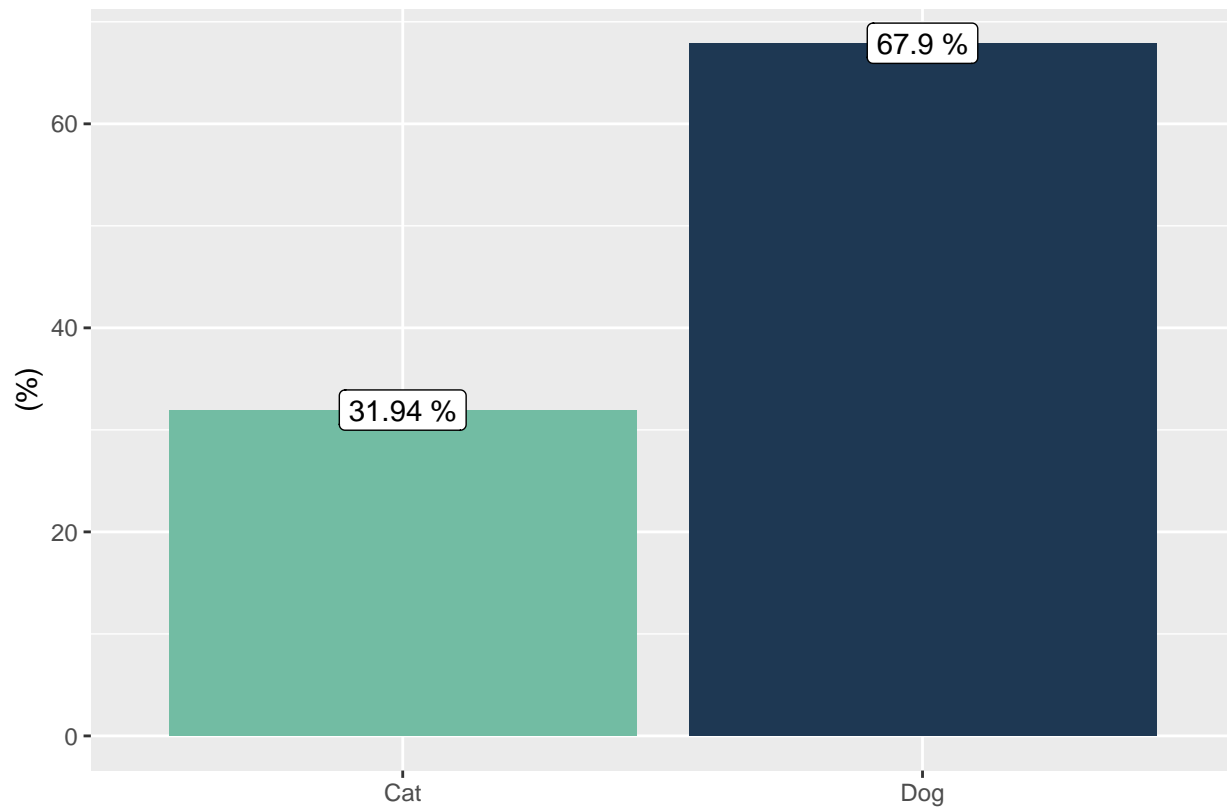
Summary Statistics

```r
general %>%
  group_by(animal) %>%
  summarise(n = n()) %>%
  mutate(n = round(100 * n / sum(n), 2)) %>%
  mutate(animal = ifelse(animal == 1, 'Dog', 'Cat')) %>%
  na.omit() %>%
  ggplot(aes(x = animal, y = n, fill = n)) +
  geom_col() +
  geom_label(aes(label = n %>% paste0(' %')), fill = 'white') +
  scale_fill_hp(guide = FALSE, begin = 0.3, end = 0.7) +
  xlab('') +
  ylab('(%)')
```

```
general %>%
  group_by(animal) %>%
  summarise(n = n()) %>%
  mutate(animal = ifelse(animal == 1, 'Dog', 'Cat')) %>%
  na.omit()
```

```
## # A tibble: 2 x 2
##   animal      n
##   <chr>   <int>
## 1 Dog      6772
## 2 Cat      3185
```

```
gg1 <- general %>%
  filter(animal == 1) %>%
  group_by(infected) %>%
  summarise(n = n()) %>%
  mutate(n = n / sum(n)) %>%
  arrange(n) %>%
  mutate(pos = (cumsum(n) - n/2)) %>%
  ggplot(aes(x = '', y = n, fill = infected)) +
  geom_col(position = 'fill') +
  coord_polar(theta ='y', start = 0) +
  geom_label(aes(label = (100*n) %>% round(2) %>% paste0('%'), x = 1, y = pos),
             fill = 'white',
             size = 5.5,
```

```r
              alpha = 0.7) +
  theme_void() +
  scale_fill_hp_d(option = 'Gryffindor2', name = 'Infected?', labels = c('No', 'Yes'), direction = -1)
  ggtitle('Dogs') +
  theme(plot.title = element_text(hjust = 0.5))

gg2 <- general %>%
  filter(animal == 2) %>%
  group_by(infected) %>%
  summarise(n = n()) %>%
  mutate(n = n / sum(n)) %>%
  arrange(n) %>%
  mutate(pos = (cumsum(n) - n/2)) %>%
  ggplot(aes(x = '', y = n, fill = infected)) +
  geom_col(position = 'fill') +
  coord_polar(theta ='y', start = 0) +
  geom_label(aes(label = (100*n) %>% round(2) %>% paste0('%'), x = 1, y = pos),
             fill = 'white',
             size = 5,
             alpha = 0.7) +
  theme_void() +
  scale_fill_hp_d(option = 'Gryffindor2', name = 'Infected?', labels = c('No', 'Yes'), direction = -1)
  ggtitle('Cats') +
  theme(plot.title = element_text(hjust = 0.5))


grid.arrange(gg1,gg2, ncol = 2)
```
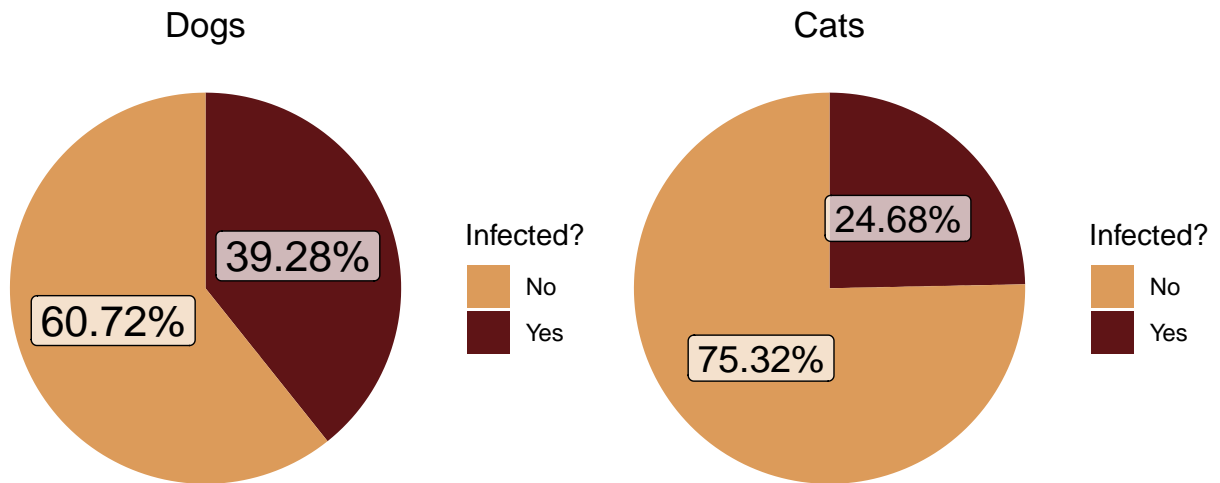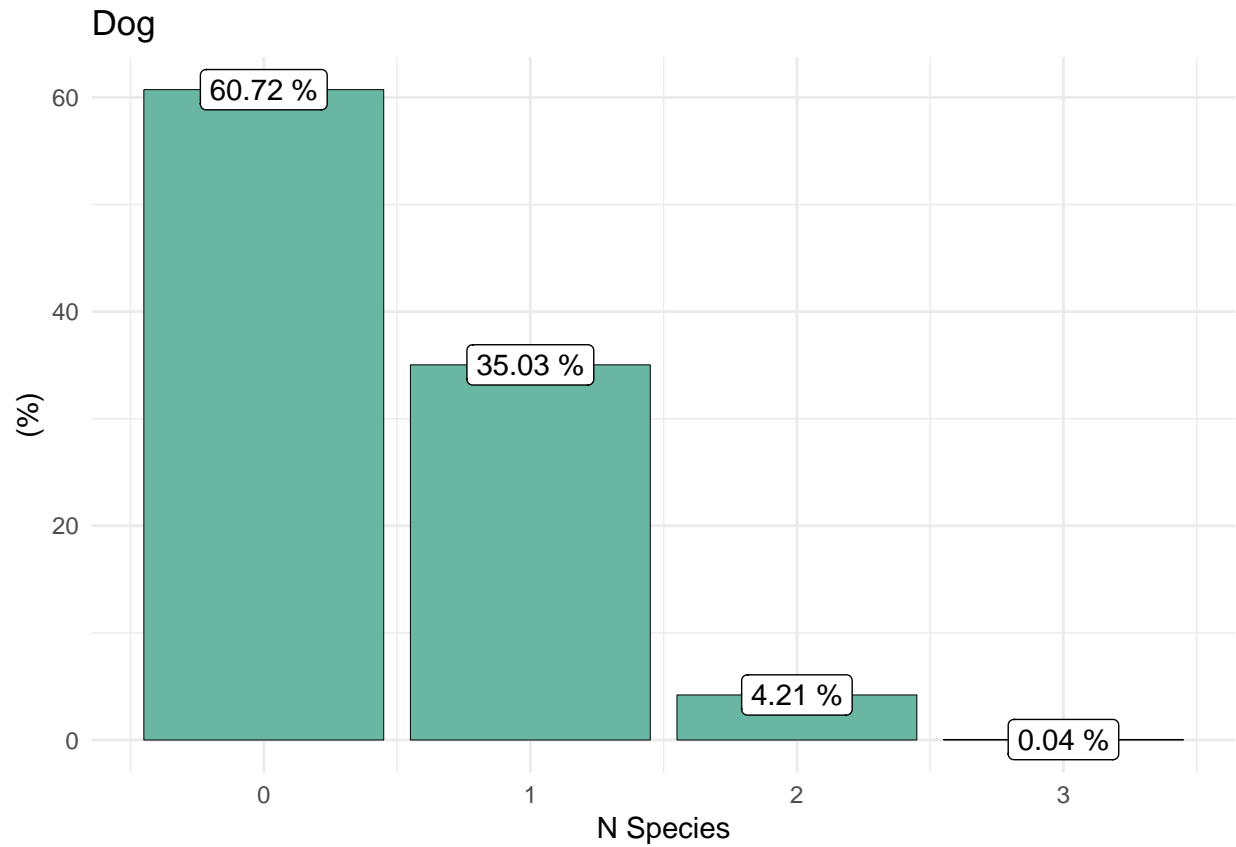
Number of Species Bacteria Frequency

```r
general %>%
  filter(animal == 1) %>%
  group_by(n_species) %>%
  summarise(n = n()) %>%
  mutate(n = 100 * n/sum(n)) %>%
  na.omit() %>%
  ggplot(aes(x = n_species, y = n)) +
  geom_col(fill = hp(10)[[4]], colour = 'black', size= 0.15) +
  geom_label(aes(label = n %>% round(2) %>% paste0(' %'))) +
  xlab('N Species') +
  ylab('(%)') +
  theme_minimal() +
  ggtitle('Dog')
```

Dog


```r
general %>%
  filter(animal == 2) %>%
  group_by(n_species) %>%
  summarise(n = n()) %>%
  mutate(n = 100 * n/sum(n)) %>%
  na.omit() %>%
  ggplot(aes(x = n_species, y = n)) +
  geom_col(fill = hp(10)[[7]], colour = 'black', size= 0.15) +
  geom_label(aes(label = n %>% round(2) %>% paste0(' %'))) +
  xlab('N Species') +
  ylab('(%)') +
  theme_minimal() +
  ggtitle('Cat')
```

Cat
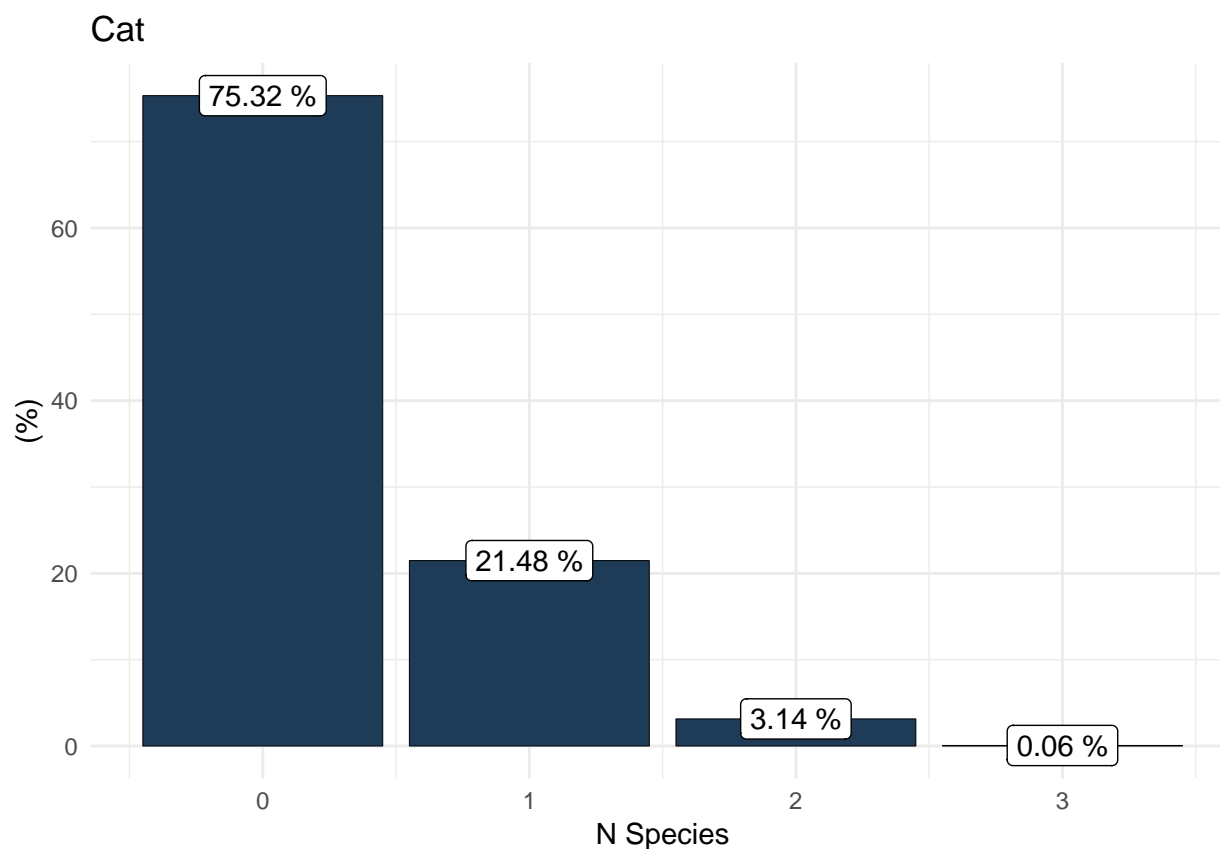
Geographic Distribution Table

province	n	infected	no_infected	proportion_infected
A Coruña	245	84	161	0.3428571
Albacete	22	9	13	0.4090909
Alicante	225	104	121	0.4622222
Almería	50	17	33	0.3400000
Araba/Álava	71	10	61	0.1408451
Asturias	318	123	195	0.3867925
Ávila	17	6	11	0.3529412
Badajoz	10	2	8	0.2000000
Balearic Islands	809	273	536	0.3374536
Barcelona	2320	764	1556	0.3293103
Bizkaia	388	131	257	0.3376289
Burgos	54	16	38	0.2962963
Cáceres	21	3	18	0.1428571
Cádiz	87	30	57	0.3448276
Cantabria	72	26	46	0.3611111
Castellón	48	18	30	0.3750000
Ceuta	22	6	16	0.2727273
Ciudad Real	26	10	16	0.3846154
Córdoba	24	9	15	0.3750000
Cuenca	28	7	21	0.2500000

province	n	infected	no_infected	proportion_infected
Gipuzkoa	93	36	57	0.3870968
Girona	76	28	48	0.3684211
Granada	32	8	24	0.2500000
Guadalajara	30	18	12	0.6000000
Huelva	7	0	7	0.0000000
Huesca	72	24	48	0.3333333
Jaén	3	0	3	0.0000000
La Rioja	37	11	26	0.2972973
Las Palmas	67	16	51	0.2388060
León	16	2	14	0.1250000
Lleida	81	39	42	0.4814815
Lugo	106	47	59	0.4433962
Madrid	2567	900	1667	0.3506038
Málaga	314	122	192	0.3885350
Melilla	10	7	3	0.7000000
Murcia	46	17	29	0.3695652
Navarre	89	36	53	0.4044944
Ourense	58	26	32	0.4482759
Palencia	5	0	5	0.0000000
Pontevedra	98	37	61	0.3775510
Salamanca	32	9	23	0.2812500
Santa Cruz de Tenerife	17	5	12	0.2941176
Segovia	4	1	3	0.2500000
Seville	130	55	75	0.4230769
Soria	2	0	2	0.0000000
Tarragona	91	24	67	0.2637363
Teruel	14	9	5	0.6428571
Toledo	43	15	28	0.3488372
Valencia	737	235	502	0.3188602
Valladolid	46	22	24	0.4782609
Zamora	3	2	1	0.6666667
Zaragoza	138	43	95	0.3115942

OGR data source with driver: ESRI Shapefile Source: "/Users/alejandro/urine_cultures_2019/data/gadm36_ESP_shp/gad:
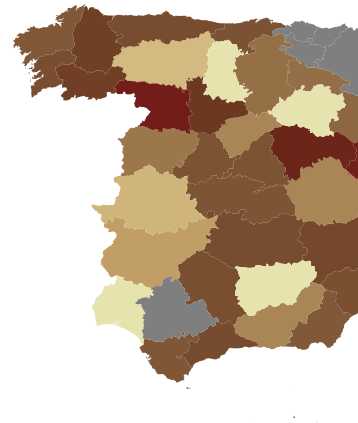
layer: "gadm36_ESP_2" with 52 features It has 13 fields

```
Pearson's Chi-squared test
```

data: . X-squared = 111.44, df = 51, p-value = 2.138e-06

Chi-Squared test outputs a p-value below 0.05. Thus it seems reasonable to *negate* that the proportion of infected population is homogeneously distributed across provinces.

Batceria Species Distribution

Number of Different Bacteria Species per Animal


Proportion of Species in Dogs
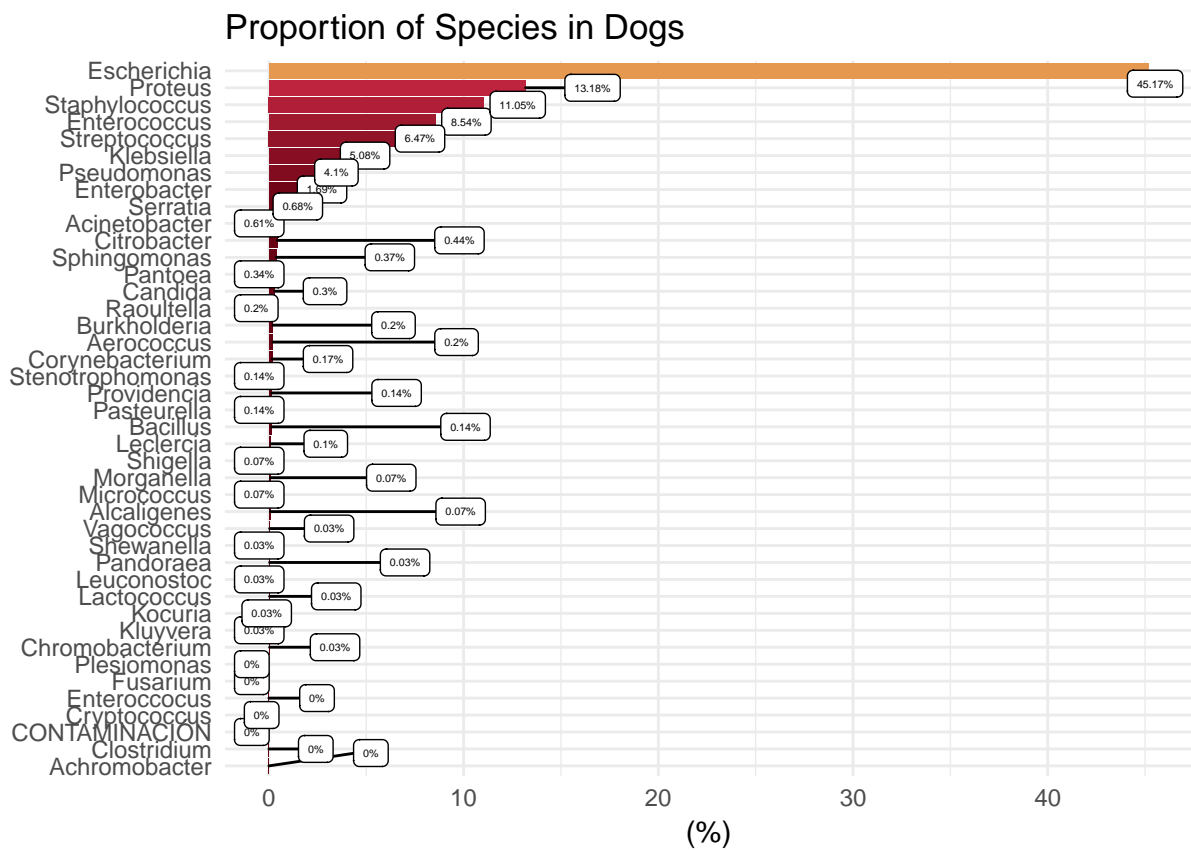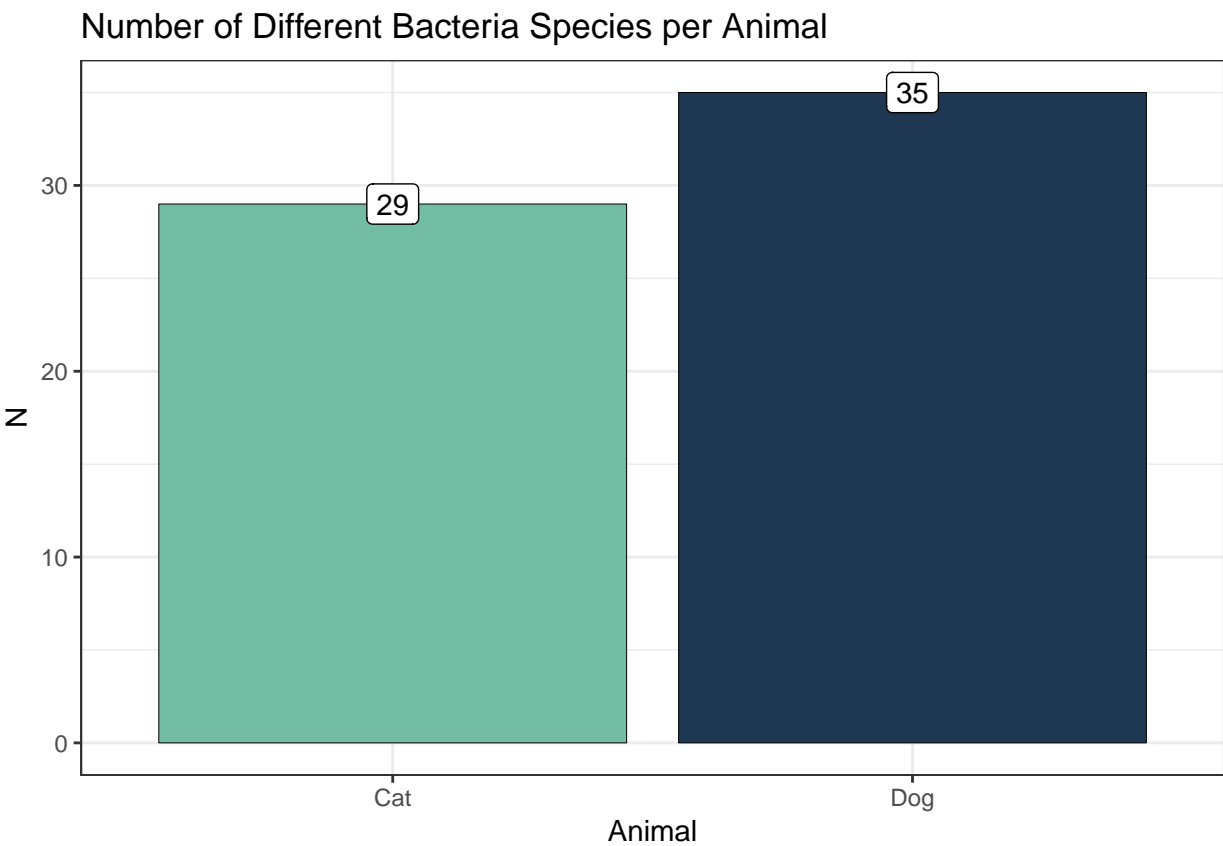
Proportion of Species in Cats
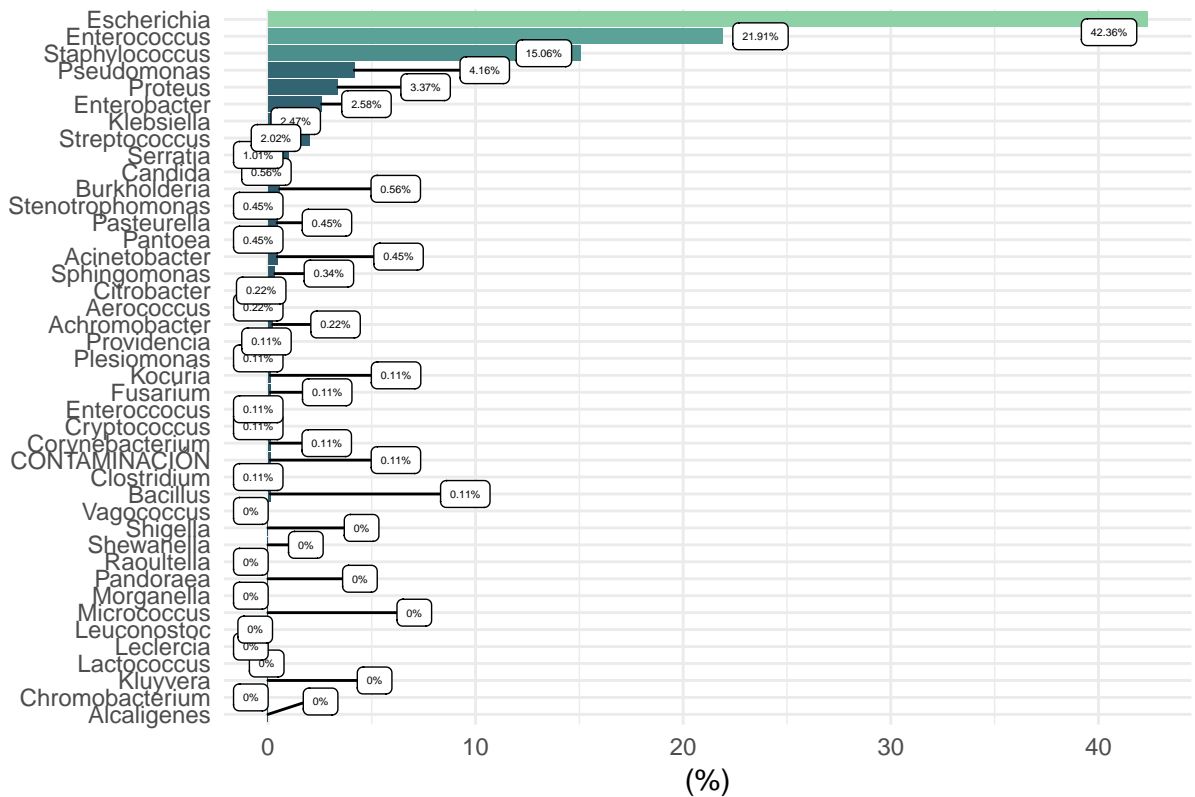
Age Difference

```
general %>%
  mutate(animal = ifelse(animal == 1, 'Dog', 'Cat')) %>%
  filter(!is.na(animal)) %>%
  ggplot(aes(x = infected, y = EDAD, fill = infected)) +
  geom_boxplot(alpha = 1, size = 0.25) +
  facet_wrap(.~animal) +
  scale_fill_hp_d(option = 'sprout', name = 'Infected?', labels = c('No', 'Yes')) +
  theme_minimal() +
  xlab('Infected?') +
  theme(axis.title.x=element_blank(),
        axis.text.x=element_blank(),
        axis.ticks.x=element_blank()) +
  ylab('Age')
```

```
## Warning: Removed 7513 rows containing non-finite values (stat_boxplot).
```
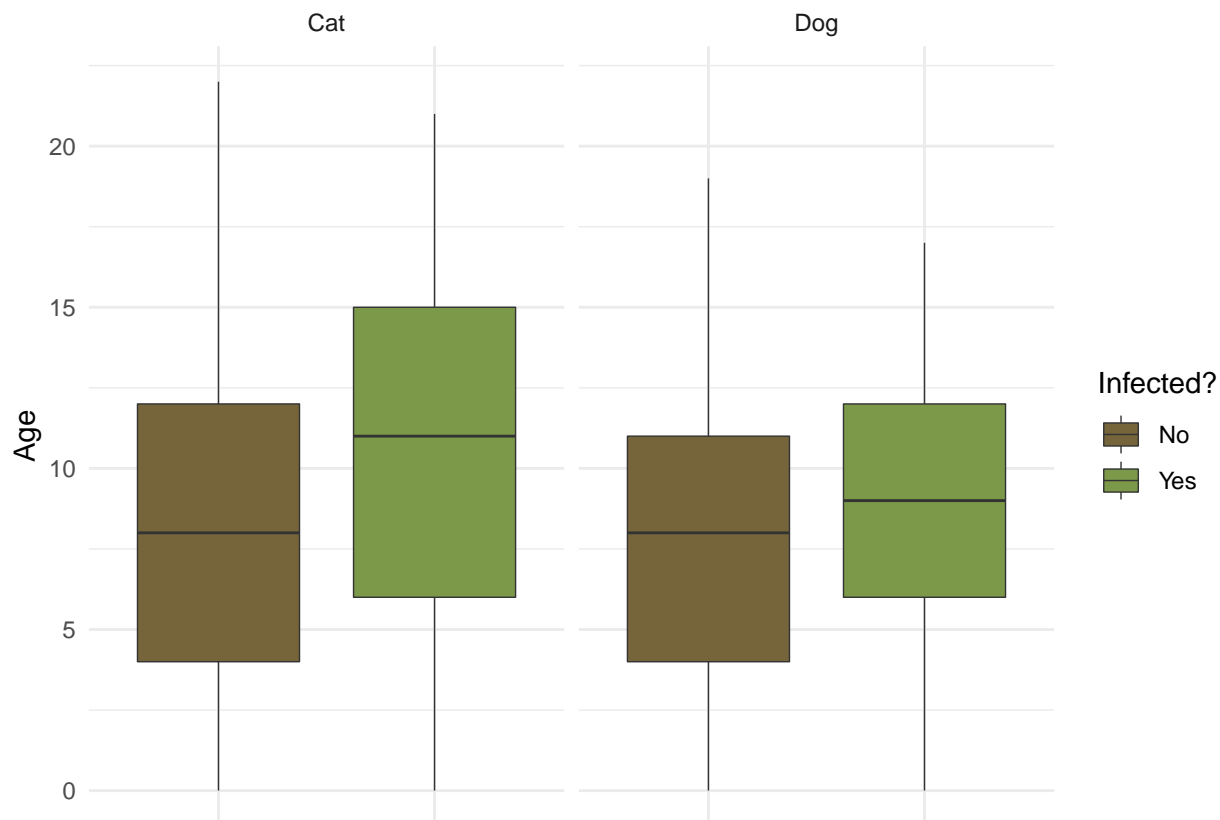
```r
general %>%
  mutate(animal = ifelse(animal == 1, 'Dog', 'Cat')) %>%
  filter(!is.na(animal)) %>%
  group_by(animal, infected) %>%
  summarise(mean_age = mean(EDAD, na.rm = TRUE),
            sd_age   = sd(EDAD,  na.rm = TRUE))
```

```
## # A tibble: 4 x 4
## # Groups:   animal [?]
##    animal infected mean_age sd_age
##    <chr>  <lgl>       <dbl>  <dbl>
## 1 Cat    FALSE        8.24   5.12
## 2 Cat    TRUE        10.5    5.30
## 3 Dog    FALSE        7.74   4.36
## 4 Dog    TRUE         8.60   4.04
```

Dogs Wilcoxon Test

```r
x <- general %>% filter(animal == 1) %>%  filter(infected) %>% .$EDAD
y <- general %>% filter(animal == 1) %>% filter(!infected) %>% .$EDAD

wilcox.test(y,x)
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  y and x
## W = 283650, p-value = 0.000184
## alternative hypothesis: true location shift is not equal to 0
```

Cat Wilcoxon Test

```r
x <- general %>% filter(animal == 2) %>%  filter(infected) %>% .$EDAD
y <- general %>% filter(animal == 2) %>% filter(!infected) %>% .$EDAD

wilcox.test(y,x)
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  y and x
## W = 45530, p-value = 1.257e-07
## alternative hypothesis: true location shift is not equal to 0
```