

Epidemiology of helminth and protozoan parasites in Dosso Region, Niger

Jacqueline Maasch & Kristen Aiemjoy

7/31/2019

Load packages: tidyverse, xtable, stargazer, ggthemes, RColorBrewer, etc.

Set working directories. Set one directory to comment depending on who is working on the file.

```
# Kristen:
#setwd("~/Dropbox/Niger parasite epi/Data/Anthro data merge")

# Jaqueline:
setwd("/Users/razel/Dropbox/Niger parasite epi/Data/")
```

Load data. Original Giardia merging was done in Stata, later this file was merged with anthropometric data from Salesforce and Schisto data. Took clean data from swab comparison manuscript. Merged script is called "GiardiaAnthroSchistoMerge.R".

```
d <- read.csv("~/Dropbox/Niger parasite epi/Data/Anthro data merge/GiardiaAnthroSchistoMerged_07.16.19.
  rowwise() %>%
  mutate(HeightM = median(Height.1, Height.2, Height.3),
         WeightM = median(Weight.1, Weight.2, Weight.3)) %>%
  ungroup() %>%
  mutate(WeightM.kg = ifelse(Weight.Unit == "Pounds", WeightM/2.205, WeightM),
         sex = as.numeric(fct_recode(gender, "1"="MALE", "2"="FEMALE")),
         standing = fct_recode(Height.Method, standing="HEIGHT", supine= "LENGTH"),
         ageD = ageM*30,
         loosewatMother = fct_collapse(sampleconsistencymother,
                                       "loose/watery" = c("Watery", "Fluffy With Ragged Edges"),
                                       "formed" = c("Lumpy", "Smooth Soft", "Separate Hard Lumps")))
```

Anthropometry.

```
#Weight for height.
d <- addWGSR(data = d, sex = "sex", firstPart = "WeightM.kg",
            secondPart = "HeightM", index = "wfh", standing="standing")

## =====
```

```
#Weight for age.
d <- addWGSR(data = d, sex = "sex", firstPart = "WeightM.kg",
            secondPart = "ageD", index = "wfa")

## =====
```

```
#Height for age.
d <- addWGSR(data = d, sex = "sex", firstPart = "HeightM",
            secondPart = "ageD", index = "hfa", standing="standing")

## =====
```

```
# svy <- svy %>%
#   mutate(ZCat = cut(wfhz, c(-Inf,-3, -2, -1, 0, Inf), labels=c("<-3", "-3,-2", "-2,-1", "-1,0", "0+")))
```

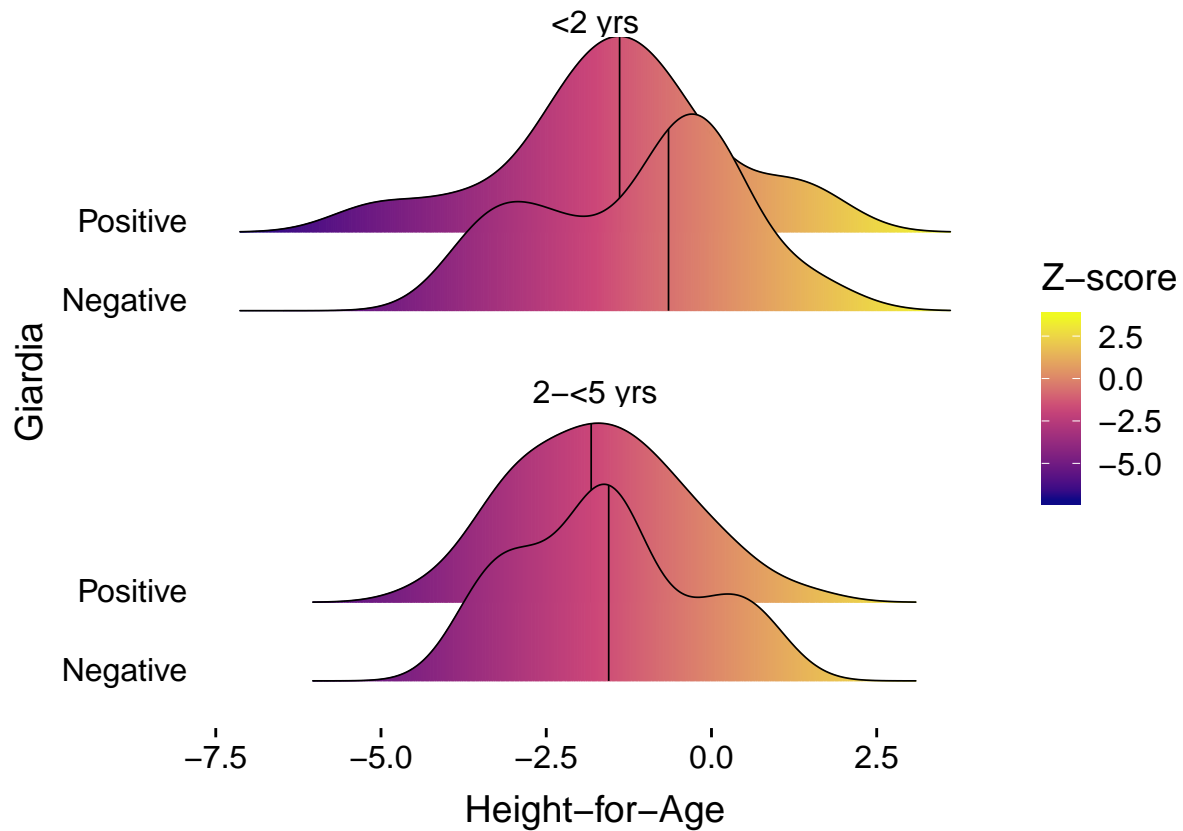
Child growth z-score results 1.

```
d <- d %>%
  mutate(AgeFactor = factor(ageY),
         AgeCat = cut(ageY, breaks= c(0, 2, 5),
                      right=FALSE, labels = c("<2 yrs", "2-<5 yrs")),
         GiardiaPos = factor(truePos, labels = c("Negative", "Positive")))

#high density infection cutoffs
d2 <- d %>%
  mutate(Above75 = as.factor(ctstoolMax >= quantile(ctstoolMax, 0.75, na.rm = T)),
         Above50 = as.factor(ctstoolMax >= quantile(ctstoolMax, 0.50, na.rm = T)))

ggplot(d, aes(x = hfaz, y = GiardiaPos)) +
  geom_density_ridges_gradient(
    #USE SCALE to change how much the distributions overlap
    aes(fill = ..x..), scale = 2.5, size = 0.3, quantile_lines = TRUE, quantiles = 2) +
  scale_fill_gradientn(
    colours = c("#0D0887FF", "#CC4678FF", "#F0F921FF"),
    name = "Z-score"
  ) +
  #theme_minimal() +
  theme_ridges(grid = FALSE, center_axis_labels = TRUE) +
  labs(x="Height-for-Age", y="Giardia") +
  #labs(x="", y="") +
  facet_wrap(~AgeCat, ncol=1) +
  theme( strip.background = element_rect(fill="white") )

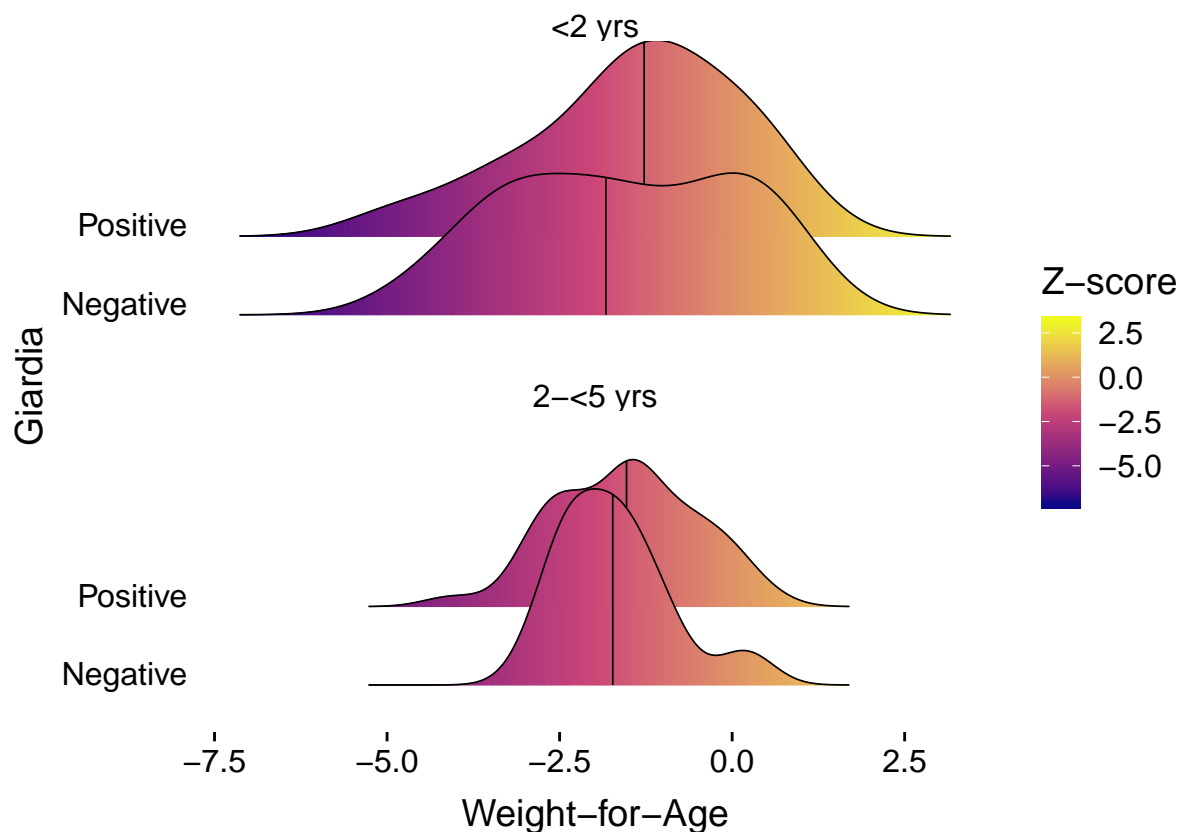
## Picking joint bandwidth of 0.681
## Picking joint bandwidth of 0.574
```



```
ggplot(d, aes(x = wfaz, y = GiardiaPos)) +
  geom_density_ridges_gradient(
    #USE SCALE to change how much the distributions overlap
    aes(fill = ..x..), scale = 2.5, size = 0.3, quantile_lines = TRUE, quantiles = 2) +
  scale_fill_gradientn(
    colours = c("#0D0887FF", "#CC4678FF", "#F0F921FF"),
    name = "Z-score"
  ) +
  #theme_minimal() +
  theme_ridges(grid = FALSE, center_axis_labels = TRUE) +
  labs(x="Weight-for-Age", y="Giardia") +
  #labs(x="", y="") +
  facet_wrap(~AgeCat, ncol=1) +
  theme( strip.background = element_rect(fill="white") )
```

```
## Picking joint bandwidth of 0.771
```

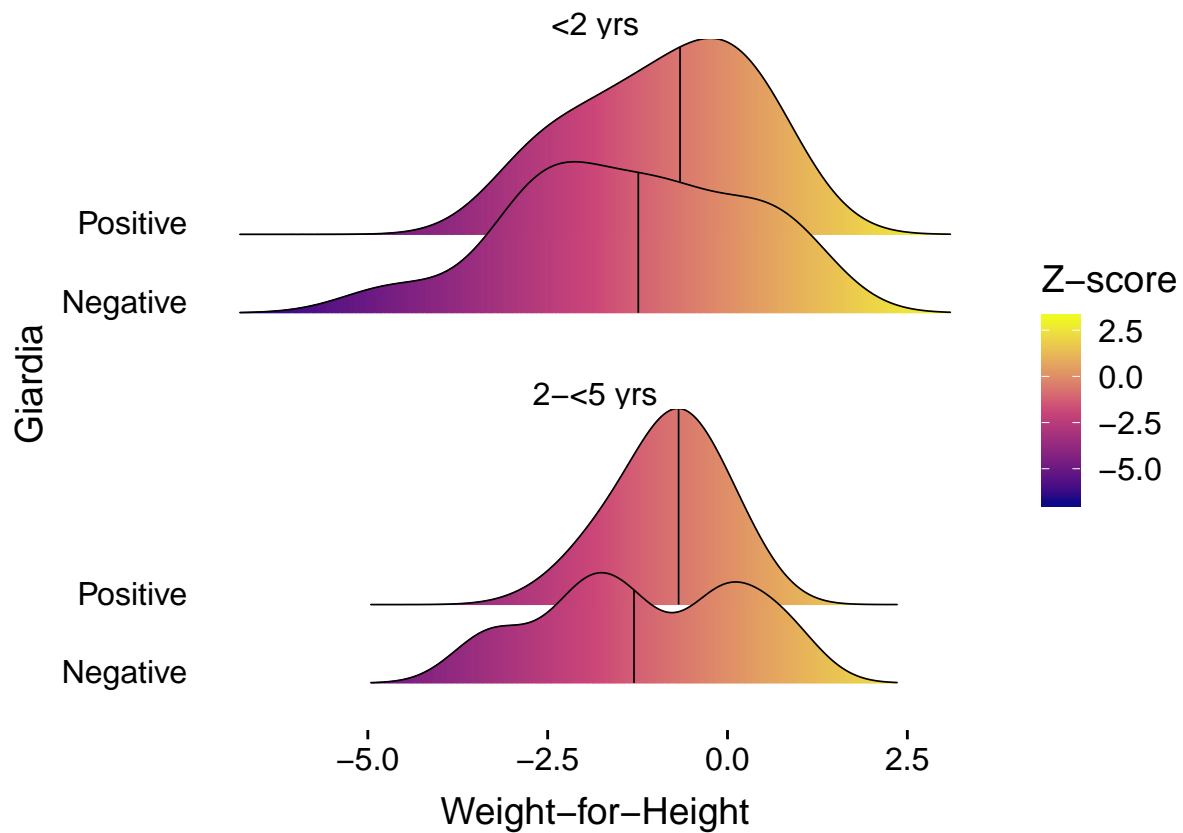
```
## Picking joint bandwidth of 0.401
```



```
ggplot(d, aes(x = wfhz, y = GiardiaPos)) +
  geom_density_ridges_gradient(
    #USE SCALE to change how much the distributions overlap
    aes(fill = ..x..), scale = 2.5, size = 0.3, quantile_lines = TRUE, quantiles = 2) +
  scale_fill_gradientn(
    colours = c("#0D0887FF", "#CC4678FF", "#F0F921FF"),
    name = "Z-score"
  ) +
  #theme_minimal() +
  theme_ridges(grid = FALSE, center_axis_labels = TRUE) +
  labs(x="Weight-for-Height", y="Giardia") +
  #labs(x="", y="") +
  facet_wrap(~AgeCat, ncol=1) +
  theme( strip.background = element_rect(fill="white") )
```

Picking joint bandwidth of 0.716

Picking joint bandwidth of 0.512



Child growth z-score results 2.

```
mean(d$hfaz)
```

```
## [1] -1.531786
```

```
sd(d$hfaz)
```

```
## [1] 1.391568
```

```
mean(d$wfaz)
```

```
## [1] -1.552976
```

```
sd(d$wfaz)
```

```
## [1] 1.218978
```

```
round(mean(d$wfaz), 2)
```

```
## [1] -0.97
```

```
round(sd(d$wfaz), 3)
```

```
## [1] 1.101
```

```
t.test(hfaz~truePos, data=d)
```

```
##
```

```
## Welch Two Sample t-test
```

```
##
```

```
## data: hfaz by truePos
```

```
## t = 0.87554, df = 48.859, p-value = 0.3856
```

```

## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -0.3760252  0.9565710
## sample estimates:
## mean in group 0 mean in group 1
## -1.334815 -1.625088

t.test(wfaz~truePos, data=d)

##
## Welch Two Sample t-test
##
## data: wfaz by truePos
## t = -0.39193, df = 45.342, p-value = 0.6969
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -0.7235035  0.4877530
## sample estimates:
## mean in group 0 mean in group 1
## -1.632963 -1.515088

t.test(wfhz~truePos, data=d)

##
## Welch Two Sample t-test
##
## data: wfhz by truePos
## t = -1.3731, df = 34.054, p-value = 0.1787
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -1.0479771  0.2028309
## sample estimates:
## mean in group 0 mean in group 1
## -1.2588889 -0.8363158

summary(lmer(hfaz ~ truePos + (1|vilID), data = d))

## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: hfaz ~ truePos + (1 | vilID)
## Data: d
##
## REML criterion at convergence: 294.4
##
## Scaled residuals:
## Min 1Q Median 3Q Max
## -2.48991 -0.73223 -0.03108 0.67563 2.31172
##
## Random effects:
## Groups Name Variance Std.Dev.
## vilID (Intercept) 0.01006 0.1003
## Residual 1.93300 1.3903
## Number of obs: 84, groups: vilID, 9
##
## Fixed effects:
## Estimate Std. Error df t value Pr(>|t|)

```

```
## (Intercept)  -1.3293      0.2703 24.4685  -4.917 4.88e-05 ***
## truePos      -0.2859      0.3261 74.9290  -0.877  0.383
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##      (Intr)
## truePos -0.814
summary(lmer(wfhz ~ truePos + (1|vilID), data = d))

## boundary (singular) fit: see ?isSingular
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: wfhz ~ truePos + (1 | vilID)
## Data: d
##
## REML criterion at convergence: 254.1
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -3.09425 -0.61777  0.07005  0.61389  2.02748
##
## Random effects:
## Groups   Name      Variance Std.Dev.
## vilID    (Intercept) 0.000     0.000
## Residual                1.187     1.089
## Number of obs: 84, groups: vilID, 9
##
## Fixed effects:
##              Estimate Std. Error    df t value Pr(>|t|)
## (Intercept)  -1.2589     0.2097 82.0000  -6.004 5.03e-08 ***
## truePos       0.4226     0.2545 82.0000   1.660  0.101
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##      (Intr)
## truePos -0.824
## convergence code: 0
## boundary (singular) fit: see ?isSingular
```

General prevalence figures.

```
smoothAgeSex <- ggplot(d, aes(x = ageM, y = as.numeric(GiardiaPos)-1)) +
  geom_smooth(aes(color = gender, fill = gender)) +
  #scale_color_viridis(discrete = TRUE, option = "D")+
  #scale_fill_viridis(discrete = TRUE) +
  theme_classic() +
  theme(legend.position = "none",
        strip.text.x = element_text(size = 12),
        axis.line = element_line(size = 0.5)) +
  scale_y_continuous(breaks = seq(0,1, .1),
                     expand = c(0,0),
                     labels = scales::percent,
```

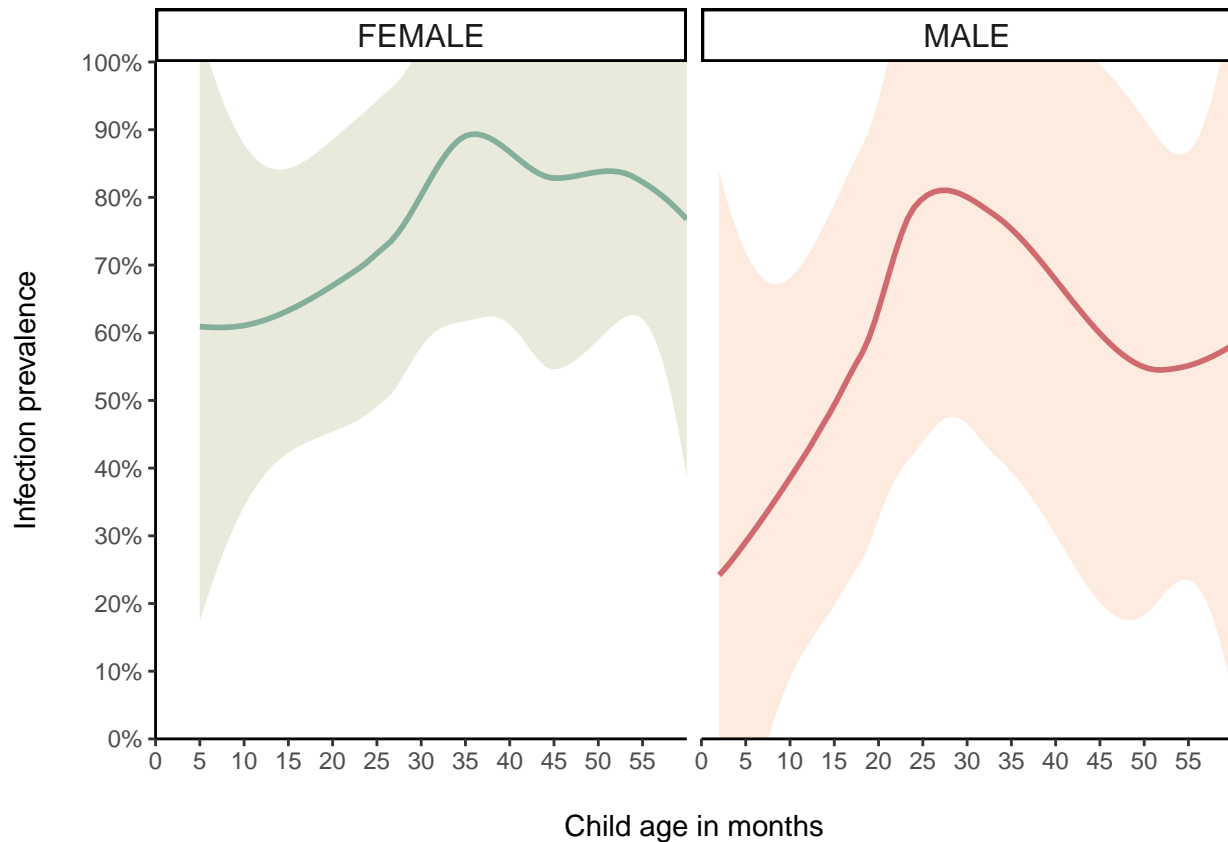
```

      minor_breaks = NULL) +
scale_x_continuous(limits = c(0,60),
  breaks = seq(0,55,5),
  expand = c(0,0),
  minor_breaks = NULL) +
scale_fill_manual(values = c("#C8C8A9", "#F9CDAD"),
  labels = c("Female", "Male")) +
scale_color_manual(name = "Specimen type",
  labels = c("Female", "Male"),
  values = c("#83AF9B", "#CF6A6E")) +
coord_cartesian(ylim = c(0,1)) +
labs(y = "Infection prevalence\n",
  x = "\nChild age in months",
  color = "", fill = "") +
facet_grid(~gender)

```

smoothAgeSex

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



```
#ggsave("/Users/razel/Dropbox/Niger parasite epi/Data/Exported figures/prevSmooth.jpg", plot = last_plot())
```

```

smoothAge <- ggplot(d, aes(x = ageM, y = as.numeric(GiardiaPos)-1)) +
  geom_smooth() +
  #scale_color_viridis(discrete = TRUE, option = "D")+
  #scale_fill_viridis(discrete = TRUE) +
  theme_minimal() +

```



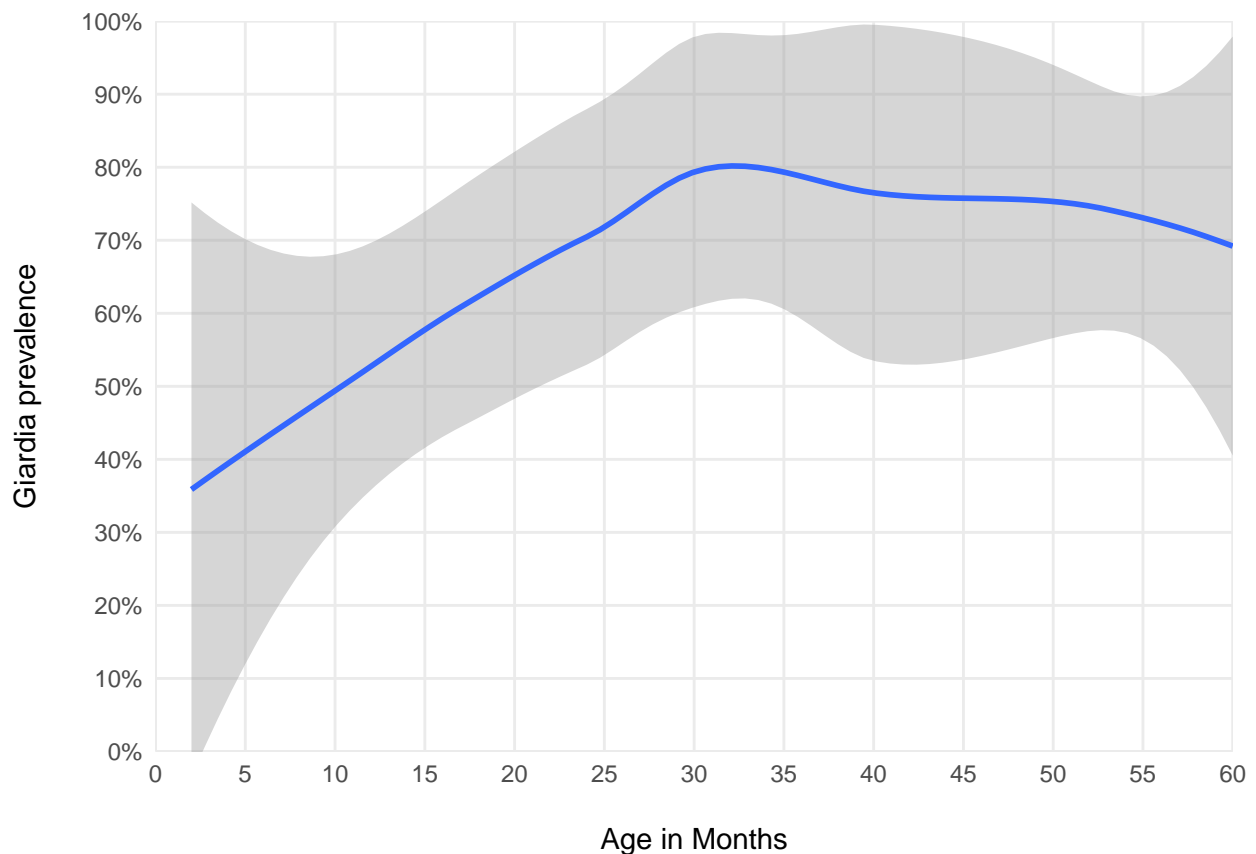
```

scale_y_continuous(breaks = seq(0,1, .1),
  expand = c(0,0),
  labels = scales::percent,
  minor_breaks = NULL) +
scale_x_continuous(limits = c(0,60),
  breaks = seq(0,60,5),
  expand = c(0,0),
  minor_breaks = NULL) +
coord_cartesian(ylim = c(0,1)) +
labs(y = "Giardia prevalence\n", x = "\nAge in Months", color = "", fill = "")

```

smoothAge

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



Parasite prevalence table.

```
#Overall.
```

```
dPrev <- read.csv("CSV files/niger_epi_prevalence_table.csv")
```

```
stargazer(dPrev, type = "latex", summary = FALSE, rownames = TRUE, column.labels = c("Species", "qPCR p
```

% Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu

% Date and time: Wed, Jul 31, 2019 - 10:51:40

```
#Disaggregating age and sex.
```

```
dPrevAge <- read.csv("CSV files/niger_epi_prevalence_age.csv")
```

Table 1: Prevalence of enteric parasites among young children in Dosso Region, Niger

	Species	qPCR.negative	qPCR.positive	Percent.positive
1	Helminths			
2	Ancylostoma duodenale	84	0	0
3	Ascaris lumbricoides	84	0	0
4	Necator americanus	84	0	0
5	Schistosoma haematobium	83	1	1.190
6	Schistosoma mansoni	84	0	0
7	Trichuris trichiura	84	0	0
8				
9	Protozoa			
10	Entamoeba histolytica	84	0	0
11	Giardia duodenalis	27	57	67.860

```
stargazer(dPrevAge, type = "latex", summary = FALSE, rownames = TRUE, column.labels = c("Species", "qPCR.negative", "qPCR.positive", "Percent.positive"))
```

% Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu

% Date and time: Wed, Jul 31, 2019 - 10:51:40

Table 2: Prevalence of enteric parasites among young children in Dosso Region, Niger

	X	Age	X.1	X.2	X.3
1		<1 year	1 year	2 years	3 years
2	Number children	11 (63.6 % female)	20 (35.0% female)	17 (64.7% female)	11 (81.8% female)
3	Helminths				
4	Ancylostoma duodenale	0 (0%)	0 (0%)	0 (0%)	0 (0%)
5	Ascaris lumbricoides	0 (0%)	0 (0%)	0 (0%)	0 (0%)
6	Necator americanus	0 (0%)	0 (0%)	0 (0%)	0 (0%)
7	Schistosoma haematobium	0 (0%)	0 (0%)	0 (0%)	0 (0%)
8	Schistosoma mansoni	0 (0%)	0 (0%)	0 (0%)	0 (0%)
9	Trichuris trichiura	0 (0%)	0 (0%)	0 (0%)	0 (0%)
10					
11	Protozoa				
12	Entamoeba histolytica	0 (0%)	0 (0%)	0 (0%)	0 (0%)
13	Giardia duodenalis	6 (54.5%)	10 (50.0%)	14 (84.4%)	9 (81.8%)

Age/prevalence sex results.

```
d %>% group_by(ageY, gender) %>%
  summarise(prev = mean(truePos),
            n = n(),
            se = sd(truePos)/sqrt(n),
            lCI = prev-1.96*se,
            uCI = prev+1.96*se)
```

```
## # A tibble: 10 x 7
## # Groups:   ageY [?]
##   ageY gender  prev     n    se    lCI    uCI
##   <int> <fct>  <dbl> <int> <dbl>  <dbl> <dbl>
## 1      0 FEMALE 0.714     7 0.184  0.353  1.08
```

```
## 2      0 MALE  0.25      4 0.25 -0.24  0.74
## 3      1 FEMALE 0.429    7 0.202  0.0326 0.825
## 4      1 MALE  0.538   13 0.144  0.256  0.821
## 5      2 FEMALE 0.818   11 0.122  0.579  1.06
## 6      2 MALE  0.833    6 0.167  0.507  1.16
## 7      3 FEMALE 0.889    9 0.111  0.671  1.11
## 8      3 MALE  0.5      2 0.5    -0.48  1.48
## 9      4 FEMALE 0.812   16 0.101  0.615  1.01
## 10     4 MALE  0.556    9 0.176  0.211  0.900
```

```
d %>% group_by(ageY) %>%
  summarise(prev = mean(truePos),
            n = n(),
            se = sd(truePos)/sqrt(n),
            lCI = prev-1.96*se,
            uCI = prev+1.96*se)
```

```
## # A tibble: 5 x 6
##   ageY prev      n      se  lCI  uCI
##   <int> <dbl> <int>  <dbl> <dbl> <dbl>
## 1     0 0.545    11 0.157  0.237 0.854
## 2     1 0.5      20 0.115  0.275 0.725
## 3     2 0.824    17 0.0953 0.637 1.01
## 4     3 0.818    11 0.122  0.579 1.06
## 5     4 0.72     25 0.0917 0.540 0.900
```

Bar charts of prevalence by age and sex.

```
dAgeSex <- d %>%
  mutate(age = as.factor(ageY)) %>%
  mutate(sex = as.factor(gender)) %>%
  select(truePos, age, sex)

dAgeSexPrev <- d %>%
  group_by(ageY) %>%
  summarise(prev = (mean(truePos))*100,
            n = n(),
            se = sd(truePos)/sqrt(n),
            lCI = prev-1.96*se,
            uCI = prev+1.96*se)

tabAgePos <- xtabs(~ truePos + age, data = dAgeSex)
tabAgePos
```

```
##           age
## truePos  0  1  2  3  4
##          0  5 10  3  2  7
##          1  6 10 14  9 18
```

```
tabAgeSex <- xtabs(~ sex + age, data = dAgeSex)
tabAgeSex
```

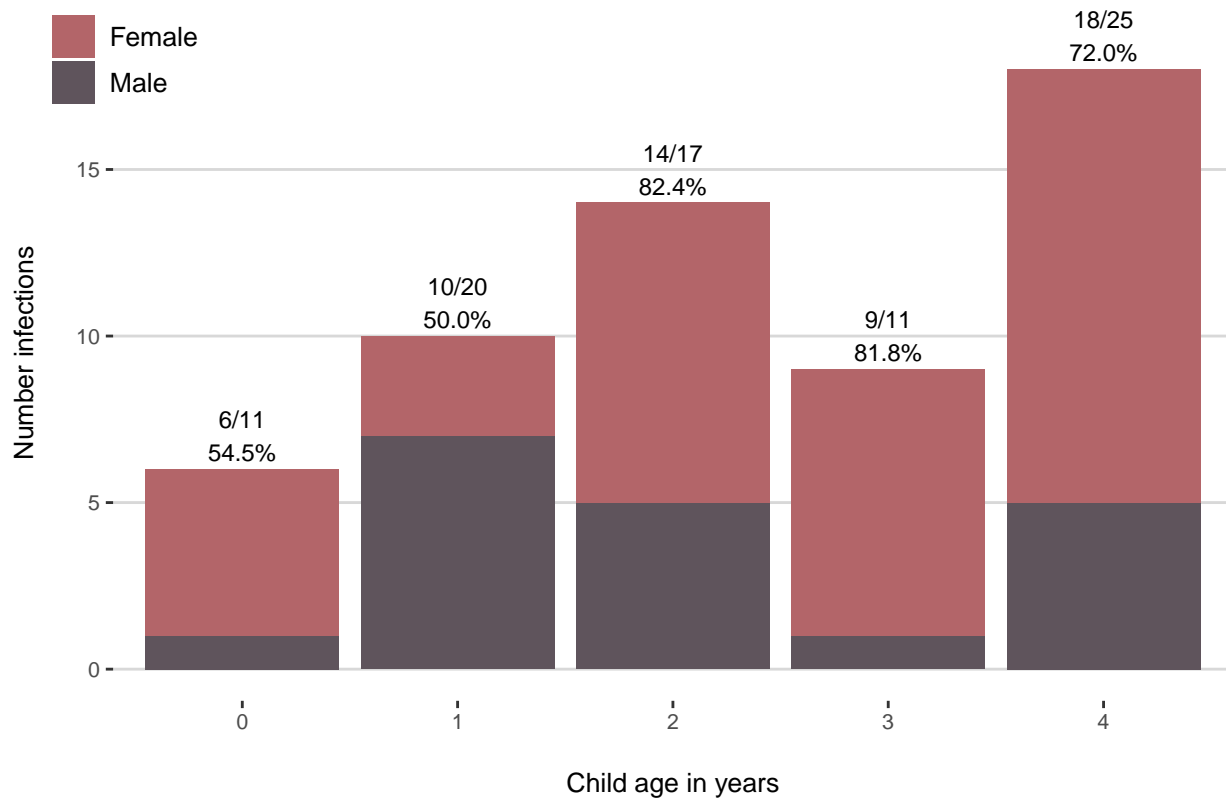
```
##           age
## sex       0  1  2  3  4
## FEMALE  7  7 11  9 16
## MALE    4 13  6  2  9
```

```

barAgeSex <- ggplot(dAgeSex, aes(x = age, y = truePos, fill = sex)) +
  geom_bar(stat = "identity") +
  theme_hc() +
  theme(axis.text = element_text(size = 8),
        axis.title = element_text(size = 10),
        plot.title = element_text(size = 10,
                                   face = "bold",
                                   hjust = 0.5),
        legend.title = element_text(size = 10),
        legend.direction = "vertical",
        legend.position = c(0.01, 0.95),
        panel.grid.major.x = element_blank()) +
  labs(y = "Number infections\n",
       x = "\nChild age in years",
       fill = "",
       title = "") +
  scale_fill_manual(values = c("#B36569", "#5F545C"),
                    label = c("Female", "Male")) +
  #geom_text(aes(label = c("6/11", "10/20", "14/17", "9/11", "18/25")))
  annotate("text", x = 1, y = 7, label = "6/11\n54.5%", size = 3) +
  annotate("text", x = 2, y = 11, label = "10/20\n50.0%", size = 3) +
  annotate("text", x = 3, y = 15, label = "14/17\n82.4%", size = 3) +
  annotate("text", x = 4, y = 10, label = "9/11\n81.8%", size = 3) +
  annotate("text", x = 5, y = 19, label = "18/25\n72.0%", size = 3)

```

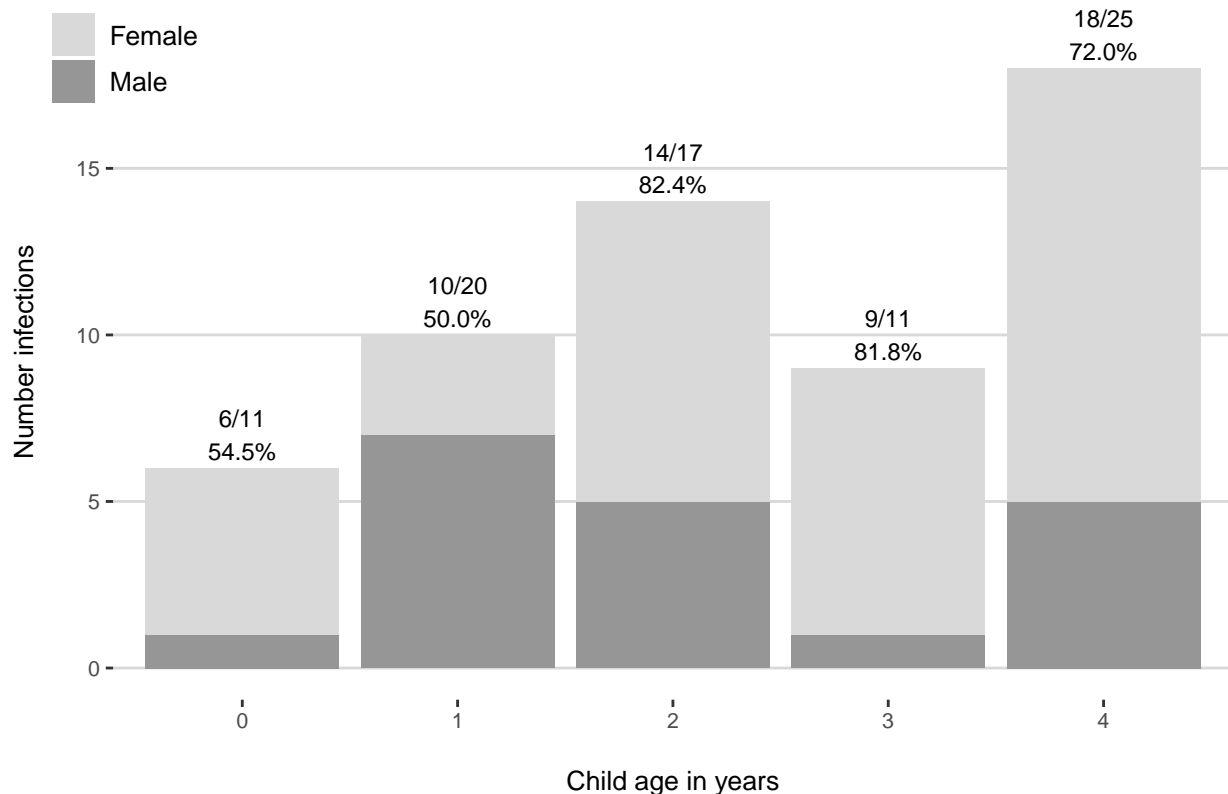
barAgeSex



```
#ggsave("/Users/razel/Dropbox/Niger parasite epi/Data/Exported figures/pos_age_sex.jpg", plot = last_pl
```

```
barAgeSexBW <- ggplot(dAgeSex, aes(x = age, y = truePos, fill = sex)) +
  geom_bar(stat = "identity") +
  theme_hc() +
  theme(axis.text = element_text(size = 8),
        axis.title = element_text(size = 10),
        plot.title = element_text(size = 10,
                                   face = "bold",
                                   hjust = 0.5),
        legend.title = element_text(size = 10),
        legend.direction = "vertical",
        legend.position = c(0.01, 0.95),
        panel.grid.major.x = element_blank()) +
  labs(y = "Number infections\n",
       x = "\nChild age in years",
       fill = "",
       title = "") +
  scale_fill_manual(values = c("#d9d9d9", "#969696"), label = c("Female", "Male")) +
  #geom_text(aes(label = c("6/11", "10/20", "14/17", "9/11", "18/25")))
  annotate("text", x = 1, y = 7, label = "6/11\n54.5%", size = 3) +
  annotate("text", x = 2, y = 11, label = "10/20\n50.0%", size = 3) +
  annotate("text", x = 3, y = 15, label = "14/17\n82.4%", size = 3) +
  annotate("text", x = 4, y = 10, label = "9/11\n81.8%", size = 3) +
  annotate("text", x = 5, y = 19, label = "18/25\n72.0%", size = 3)
```

barAgeSexBW



```
#ggsave("/Users/razel/Dropbox/Niger parasite epi/Data/Exported figures/pos_age_sex_BW.jpg", plot = last.
```

Stool consistency by the modified Bristol Stool Form Scale for children (mBSFS-C).

```
addmargins(table(d$truePos, d$loosewat1))
```

```
##
##      formed loose/watery Sum
##    0      22           5  27
##    1      51           6  57
##   Sum     73          11  84
```

```
addmargins(table(d$truePos, d$loosewatMother))
```

```
##
##      loose/watery formed Sum
##    0          12      15  27
##    1          25      32  57
##   Sum          37      47  84
```

```
dDiar <- d %>%
  select(personID, truePos, loosewat1, loosewatMother)
```

```
#11/84 children had loose/watery stool = 13.1%
#73/84 children had formed stool = 86.9%
#37/84 mothers had loose watery stool = 44.0%
#47/84 mothers had formed stool = 56.0%
```

```
tabDiarFam <- xtabs(~loosewat1 + loosewatMother, data = dDiar)
tabDiarFam
```

```
##              loosewatMother
## loosewat1      loose/watery formed
##   formed              27      46
##   loose/watery         10      1
```

```
#46/84 child:formed/mother:formed = 54.8%
#27/84 child:formed/mother:loose-watery = 32.1%
#10/84 child:loose-watery/mother:loose-watery = 11.9%
#1/84 child:loose-watery/mother:formed = 1.2%
```

```
#child:mother agree = 66.7%
#child:mother disagree = 33.3%
```

```
tabDiarPos <- xtabs(~ truePos+ loosewat1, data = dDiar)
tabDiarPos
```

```
##      loosewat1
## truePos formed loose/watery
##    0      22           5
##    1      51           6
```

```
#51/57 Giardia positives had formed stool = 89.5%
#32/51 of their mothers had formed stool = 62.7%
#6/57 Giardia positives had loose/watery stool = 10.5%
#6/6 of their mothers had loose/watery stool = 100%
```

Study population characteristics.

```
#All villages.
levels(d$vilID)
```

```
[1] "BABA DEY" "GOBERI PEULH" "GUILLARE PEULH"
[4] "LISSORE" "MOUNBEINA FANDOGA" "POULLO"
[7] "SETTI I" "TOMBO" "WERE DJATAME PEUL"
```

```
#Mean age in months.
summary(d$ageM)
```

```
Min. 1st Qu. Median Mean 3rd Qu. Max. 2.00 17.75 33.00 33.55 53.25 60.00
```

```
#Mean/median age in years.
summary(d$ageY)
```

```
Min. 1st Qu. Median Mean 3rd Qu. Max. 0.000 1.000 2.000 2.226 4.000 4.000
```

```
#Number male and female.
table(d$gender)
```

```
FEMALE MALE 50 34
```

```
#Giardia prevalence.
Table1_a <- d %>%
  group_by(truePos) %>%
  summarize(nChildren = n(),
    nVil = length(unique(vilID)),
    MedianAge = round(median(ageY, na.rm = T), 2),
    Age.p25 = round(quantile(ageY, na.rm = T, probs=.25), 2),
    Age.p75 = round(quantile(ageY, na.rm = T, probs=.75), 2),
    prop.female=round(100*mean(gender == "FEMALE", na.rm=TRUE),1),
    nFemale = sum(gender=="FEMALE", na.rm=TRUE)) %>%
  mutate(
    Female = paste(nFemale, " (",prop.female,"% )", sep=""),
    #N_Vil = paste (nChildren, " (",nPools , ") ", sep=""),
    MedianAge = paste(MedianAge, " (", Age.p25, "-", Age.p75 , ") ", sep="")) %>%
  select(nChildren, nVil, MedianAge, Female) %>%
  t %>% #transpose
  as.data.frame() %>%
  `colnames<-`(c("Negative", "Positive"))
#select(Watery, Loose, Smooth, Lumpy, Pellets)
```

```
Table1_b <- d %>%
  summarize(nChildren = n(),
    nVil = length(unique(vilID)),
    MedianAge = round(median(ageY, na.rm = T), 2),
    Age.p25 = round(quantile(ageY, na.rm = T, probs=.25), 2),
    Age.p75 = round(quantile(ageY, na.rm = T, probs=.75), 2),
    prop.female=round(100*mean(gender == "FEMALE", na.rm=TRUE),1),
    nFemale = sum(gender=="FEMALE", na.rm=TRUE)) %>%
  mutate(
    Female = paste(nFemale, " (",prop.female,"% )", sep=""),
    MedianAge = paste(MedianAge, " (", Age.p25, "-", Age.p75 , ") ", sep="")) %>%
  select(nChildren, nVil, MedianAge, Female) %>%
  t %>% #transpose
  as.data.frame() %>%
  `colnames<-`(c("Overall"))
```

```
Table1 <- cbind(Table1_a, Table1_b) %>%
  select(Negative, Positive, Overall)
row.names(Table1) <- c("N() Children", "N() Communities", "Median age (IQR)", "N(%) female")
colnames(Table1) <- c("PCR negative (Stool)", "PCR positive (Stool)", "Overeall")
```

Table1

	PCR negative (Stool)	PCR positive (Stool)	Overeall
N() Children	27 57 84	N() Communities 9 7 9	Median age (IQR) 1 (1-3.5) 2 (1-4) 2 (1-4)
N(%) female	12 (44.4%)	38 (66.7%)	50 (59.5%)

#Table1 in stargazer

```
stargazer(Table1, type = "latex", summary = FALSE, rownames = FALSE, column.labels = c("Negative", "Posi
```

% Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu
 % Date and time: Wed, Jul 31, 2019 - 10:51:43

Table 3: Population summary by detection of *Giardia intesinalis* in bulk stool

PCR negative (Stool)	PCR positive (Stool)	Overeall
27	57	84
9	7	9
1 (1-3.5)	2 (1-4)	2 (1-4)
12 (44.4%)	38 (66.7%)	50 (59.5%)

#Table1 in stargazer -- transposed

```
Table1T <- t(Table1)
```

```
stargazer(Table1T, type = "latex", summary = FALSE, rownames = FALSE, column.labels = c("Negative", "Posi
```

% Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu
 % Date and time: Wed, Jul 31, 2019 - 10:51:43

Table 4: Population summary by detection of *Giardia intesinalis* in bulk stool

N() Children	N() Communities	Median age (IQR)	N(%) female
27	9	1 (1-3.5)	12 (44.4%)
57	7	2 (1-4)	38 (66.7%)
84	9	2 (1-4)	50 (59.5%)

Giardia prevalence by village.

```
d %>% group_by(vilID) %>%
  summarise(prev = mean(truePos),
            n = n(),
            se = sd(truePos)/sqrt(n),
            lCI = prev-1.96*se,
            uCI = prev+1.96*se)
```

A tibble: 9 x 6

##	vilID	prev	n	se	lCI	uCI
##	<fct>	<dbl>	<int>	<dbl>	<dbl>	<dbl>
## 1	BABA DEY	0.778	18	0.101	0.580	0.975


```
## 2 GOBERI PEULH      0      1 NA      NA      NA
## 3 GUILLARE PEULH    0.722    18 0.109    0.509    0.935
## 4 LISSORE           0.5      2 0.5     -0.48    1.48
## 5 MOUNBEINA FANDOGA 0.875    16 0.0854   0.708    1.04
## 6 POULLO            0.692    13 0.133    0.431    0.953
## 7 SETTI I           0.333     9 0.167    0.00667  0.660
## 8 TOMBO             0       3 0       0       0
## 9 WERE DJATAME PEUL 0.75     4 0.25    0.26    1.24
```

```
#####
```

```
addmargins(table(d$vilID, d$truePos))
```

```
##
##           0  1 Sum
## BABA DEY    4 14 18
## GOBERI PEULH 1  0  1
## GUILLARE PEULH 5 13 18
## LISSORE      1  1  2
## MOUNBEINA FANDOGA 2 14 16
## POULLO       4  9 13
## SETTI I      6  3  9
## TOMBO        3  0  3
## WERE DJATAME PEUL 1  3  4
## Sum         27 57 84
```

```
CrossTable(d$vilID, d$truePos, prop.r = T, prop.c = F, prop.t = F, prop.chisq = F)
```

```
##
##
##      Cell Contents
## |-----|
## |                      N |
## |          N / Row Total |
## |-----|
##
##
## Total Observations in Table:  84
```

```
##
##           | d$truePos
##           | 0 | 1 | Row Total |
## -----|-----|-----|-----|
## BABA DEY | 4 | 14 | 18 |
##           | 0.222 | 0.778 | 0.214 |
## -----|-----|-----|-----|
## GOBERI PEULH | 1 | 0 | 1 |
##           | 1.000 | 0.000 | 0.012 |
## -----|-----|-----|-----|
## GUILLARE PEULH | 5 | 13 | 18 |
##           | 0.278 | 0.722 | 0.214 |
## -----|-----|-----|-----|
## LISSORE | 1 | 1 | 2 |
##           | 0.500 | 0.500 | 0.024 |
## -----|-----|-----|-----|
```

```
## MOUNBEINA FANDOGA |      2 |      14 |      16 |
##                   |    0.125 |    0.875 |    0.190 |
## -----|-----|-----|-----|
##           POULLO |      4 |      9 |      13 |
##                   |    0.308 |    0.692 |    0.155 |
## -----|-----|-----|-----|
##           SETTI I |      6 |      3 |      9 |
##                   |    0.667 |    0.333 |    0.107 |
## -----|-----|-----|-----|
##           TOMBO |      3 |      0 |      3 |
##                   |    1.000 |    0.000 |    0.036 |
## -----|-----|-----|-----|
## WERE DJATAME PEUL |      1 |      3 |      4 |
##                   |    0.250 |    0.750 |    0.048 |
## -----|-----|-----|-----|
##           Column Total |      27 |      57 |      84 |
## -----|-----|-----|-----|
##
##
```

```
posVil <- as.data.frame(addmargins(table(d$vilID, d$truePos)))
vilNeg <- posVil[1:10, c(1, 3)]
vilPos <- posVil[11:20, c(1, 3)]
vilSum <- posVil[21:30, c(1, 3)]
vilPosNeg <- left_join(vilNeg, vilPos, by = "Var1")
vilPosNegSum <- left_join(vilPosNeg, vilSum, by = "Var1")
colnames(vilPosNegSum) <- c("Village", "Negative", "Positive", "Total")
vilPosNegSum = mutate(vilPosNegSum,
                      percentPos = (Positive / Total)*100)
colnames(vilPosNegSum) <- c("Village", "Negative", "Positive", "Total", "Percent Positive")
vilPosNegSum <- vilPosNegSum[,c(1,2,3,5,4)]

stargazer(vilPosNegSum, type = "latex", summary = FALSE, rownames = FALSE,
          title = "Real-time PCR detection of \textit{Giardia duodenalis} by village")
```

```
##
## % Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu
## % Date and time: Wed, Jul 31, 2019 - 10:51:43
## \begin{table}[!htbp] \centering
##   \caption{Real-time PCR detection of \textit{Giardia duodenalis} by village}
##   \label{}
##   \begin{tabular}{@{\extracolsep{5pt}} ccccc}
##     \hline
##     Village & Negative & Positive & Percent Positive & Total \\
##     \hline
##     BABA DEY & 2 & 14 & 87.778 & 16 \\
##     GOBERI PEULH & 4 & 9 & 22.222 & 13 \\
##     GUILLARE PEULH & 5 & 3 & 60.000 & 8 \\
##     LISSORE & 1 & 3 & 30.000 & 4 \\
##     MOUNBEINA FANDOGA & 2 & 14 & 87.500 & 16 \\
##     POULLO & 4 & 9 & 69.231 & 13 \\
##     SETTI I & 6 & 3 & 33.333 & 9 \\
##     TOMBO & 3 & 0 & 0.000 & 3 \\
##     WERE DJATAME PEUL & 1 & 3 & 75.000 & 4 \\
##   \end{tabular}
```

```
## Sum & $27$ & $57$ & $67.857$ & $84$ \\  
## \hline \)[-1.8ex]  
## \end{tabular}  
## \end{table}
```

Schisto positive individual.

```
dSchisto <- d %>%  
  filter(St1_schisto_ct != 19.41028)
```