

# BombSurv2015

*Samantha Alger & Alex Burnham*

*October 27, 2016*

## Metadata

**Author:** Samantha Alger & Alex Burnham

**Date:** 27 October 2016

**Data Set:** These data were collected during the a 2015 Bumble Bee Survey in Vermont by Samantha Alger and Alex Burnham with all testing is being done at the University of Vermont

**Data Source:** 2015 Bumble Bee Survey

**Funding Source:** Garden Club of America, Pollinator Partnership, Fabri Fiahlo (UVM), Roger Williams Park Zoo Sophie Danforth Conservation Fund.

**Data Collection:** Bumble bees were netted from flowers at field sites.

**Columns:** (from left to right) Site ID, Target (Actin, DWV, IAPV, or BQCV), sample ID, From the qPCR results: Ctmean, Ctsd, quantity mean, quantitysd, Run # (QPCR), Date specimen was collected, Field ID, Bee species, plant the bee was caught on, date the specimen was processed in the lab, was pollen found on the leg and collected (0-no, 1- yes), dil.factor (RNA diluted), apiary present/not present at site (1, 0), genome copy for target (not normalized), ACTIN genome copy, normalized target genome copy (normalized using ACTIN), 6-10 Honey Bees were collected at each site, homogenized and assayed for viruses, 'norm\_genome\_copbeeHB' is the normalized genome copy for honey bee for the target specified, 'CTmeanhb' is the CT value for the honey bee for each target (from qpcr data), 'HB collected' is whether honeybees were collected and processed at that site, 1 or NA, virus 'BINYPrefilter' is whether the bumble bee was positive for the virus of interest, 'virusBINY' is whether the bumble bee was positive for the virus but takes into account the limit of detection for each target, 'HB\_Abun' is # of honey bees over total number bee observed- taken from bee abundance survey at each site. 'HBSiteBin' is for DWV data only. Honey Bee virus load results showed two distinct groups 'high' (VL) and 'low' (VL).

**Rows:** Bumble bee specimens

**Missing values:** NA

---

```
#Preliminaries:
```

```
# Clear memory of characters
```

```
ls()
```

```
## character(0)
```

```
rm(list=ls())
```

```
# Call blue color palette for graphics
```

```
library(RColorBrewer)
```

```
library(ggplot2)
```

```
library(dplyr)
```

```
##
```

```
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
```

```
##
```

```

##      filter, lag
## The following objects are masked from 'package:base':
##
##      intersect, setdiff, setequal, union
library(plyr)

## -----
## You have loaded plyr after dplyr - this is likely to cause problems.
## If you need functions from both plyr and dplyr, please load plyr first, then dplyr:
## library(plyr); library(dplyr)
## -----
##
## Attaching package: 'plyr'
## The following objects are masked from 'package:dplyr':
##
##      arrange, count, desc, failwith, id, mutate, rename, summarise,
##      summarize
# Set Working Directory
setwd("~/AlgerCollaborations/2015_Bombus_Survey/CSV_Files")

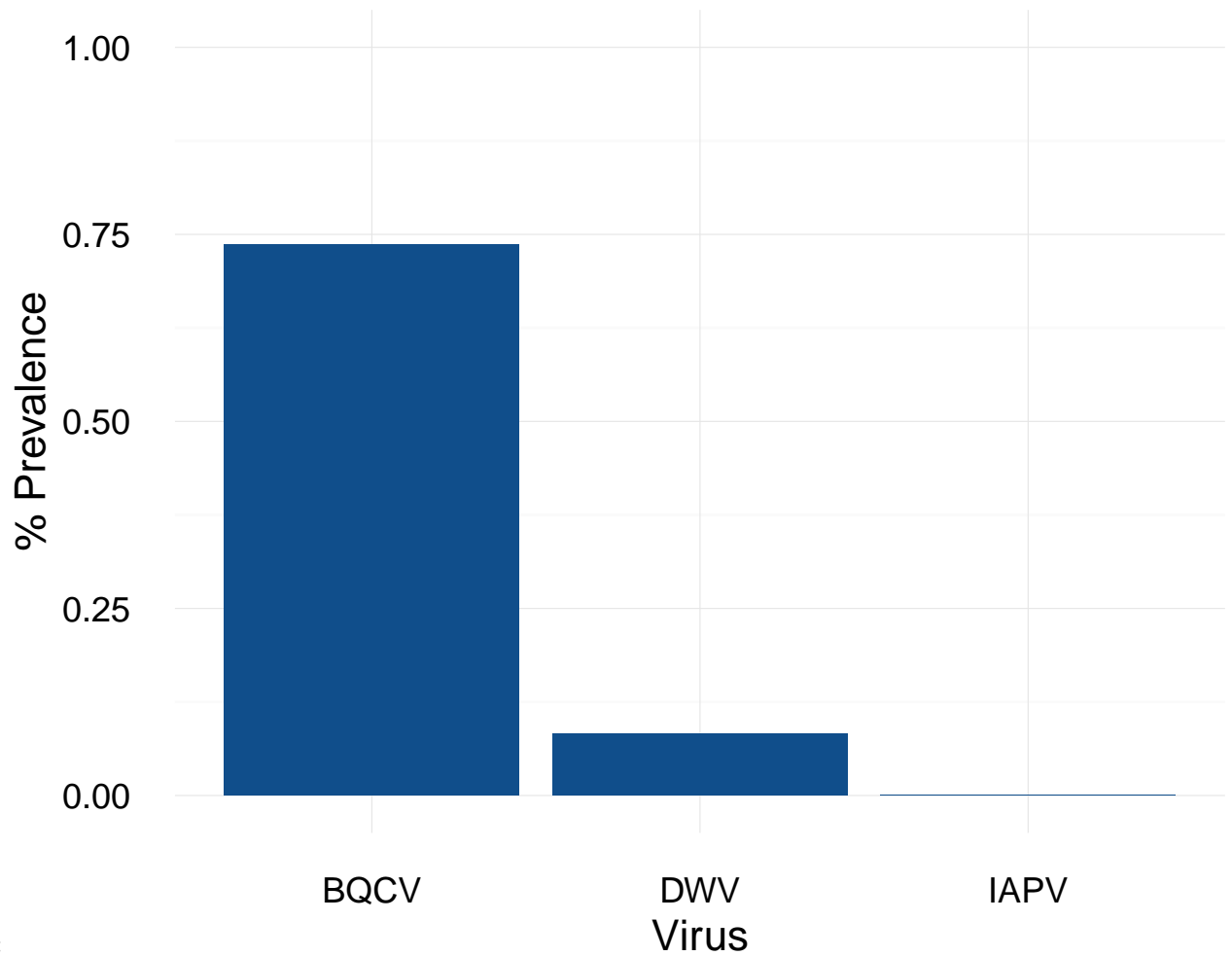
# read in data
BombSurv <- read.table("BombSurv.csv",header=TRUE,sep="," ,stringsAsFactors=FALSE)

BombSurv2 <- read.table("BombSurv2.csv",header=TRUE,sep="," ,stringsAsFactors=FALSE)

# remove unneeded columns from the DF
BombSurv <- select(BombSurv, -X, -Ct_mean, -Ct_sd, -quantity_mean, -quantity_sd, -run, -date_processed,

# remove unneeded columns from the DF
BombSurv2 <- select(BombSurv2, -City, -Name, -virusBINY_PreFilter)

```



Prevelence:

Prevalence by Species:

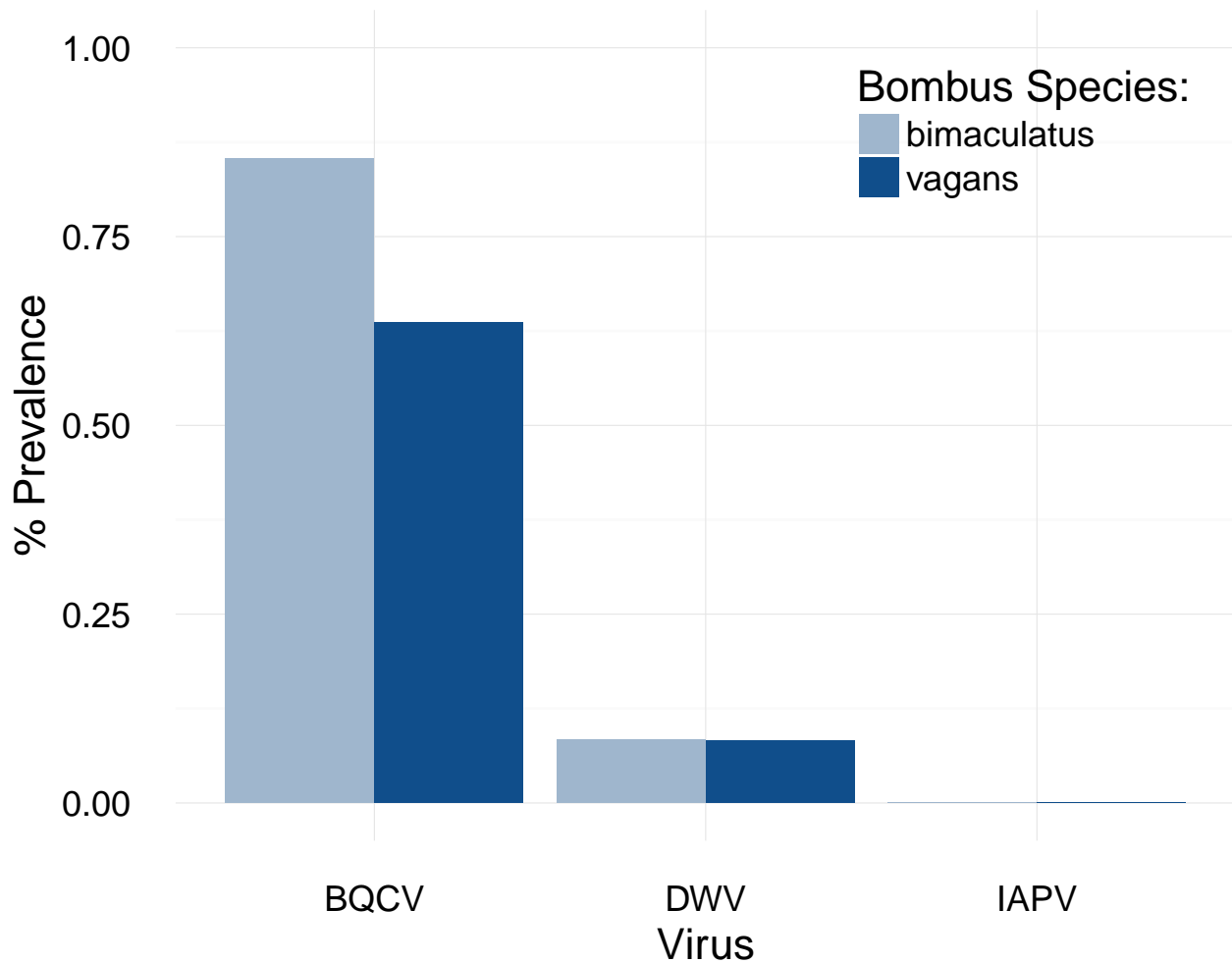
```
# using dplyr to get summary of virusBINY by species and target name:
VirusSum <- dplyr::summarise(BombSurv, c("species", "target_name"), summarise,
  n = length(virusBINY),
  mean = mean(virusBINY, na.rm=TRUE),
  sd = sd(virusBINY, na.rm=TRUE),
  se = sd / sqrt(n))
```

```
# remove data for species other than vagans and :
VirusSum <-VirusSum[-c(4:9),]
```

```
#choosing color pallet
colors <- c("slategray3", "dodgerblue4")
```

```
#Create a bar graph for viruses by bombus species (aes= aesthetics):
plot1 <- ggplot(VirusSum, aes(x=target_name, y=mean, fill=species)) +
  geom_bar(stat="identity",
    position=position_dodge()) + labs(x="Virus", y = "% Prevalence")
```

```
plot1 + theme_minimal(base_size = 17) + scale_fill_manual(values=colors, name="Bombus Species:", labels=)
```



**Figure 1:** Percent prevalence (% of infected individuals) for two bumble bee species, *Bombus bimaculatus* and *Bombus vagans* for black queen cell virus (BQCV), deformed wing virus (DWV) and Israeli acute paralysis virus (IAPV). BQCV was significantly more prevalent among *B. bimaculatus* than *B. vagans* ( $p < 0.00001$ ). No significant differences in DWV prevalence between species ( $p = 0.9469$ ). IAPV was detected only in low levels among a few individuals. This figure only includes bees with quantifiable viral infections above the threshold of detection.

```
#stats for % prevalence, species differences
statsplit <- split(BombSurv, BombSurv$target_name)

# chi.sq test for BQCV prev. vs species (p < 0.00001)
chisq.test(statsplit$BQCV$virusBINY, statsplit$BQCV$species)

## Warning in chisq.test(statsplit$BQCV$virusBINY, statsplit$BQCV$species):
## Chi-squared approximation may be incorrect

##
## Pearson's Chi-squared test
##
## data: statsplit$BQCV$virusBINY and statsplit$BQCV$species
## X-squared = 26.846, df = 3, p-value = 6.343e-06
fisher.test(statsplit$BQCV$virusBINY, statsplit$BQCV$species)

##
```

```

## Fisher's Exact Test for Count Data
##
## data: statsplit$BQCV$virusBINY and statsplit$BQCV$species
## p-value = 1.803e-06
## alternative hypothesis: two.sided

# chi.sq test for DWV prev. vs species (p = 0.9469)
chisq.test(statsplit$DWV$virusBINY, statsplit$DWV$species)

## Warning in chisq.test(statsplit$DWV$virusBINY, statsplit$DWV$species): Chi-
## squared approximation may be incorrect

##
## Pearson's Chi-squared test
##
## data: statsplit$DWV$virusBINY and statsplit$DWV$species
## X-squared = 0.36724, df = 3, p-value = 0.9469

# make apiary binary a character so we can use it as a factor in the model
BombSurv$apiary_near_far <- as.character(BombSurv$apiary_near_far)

# using ddply to get summary of virusBINY by species and target name:
VirusSum1 <- ddply(BombSurv, c("apiary_near_far", "target_name"), summarise,
  n = length(virusBINY),
  mean = mean(virusBINY, na.rm=TRUE),
  sd = sd(virusBINY, na.rm=TRUE),
  se = sd / sqrt(n))

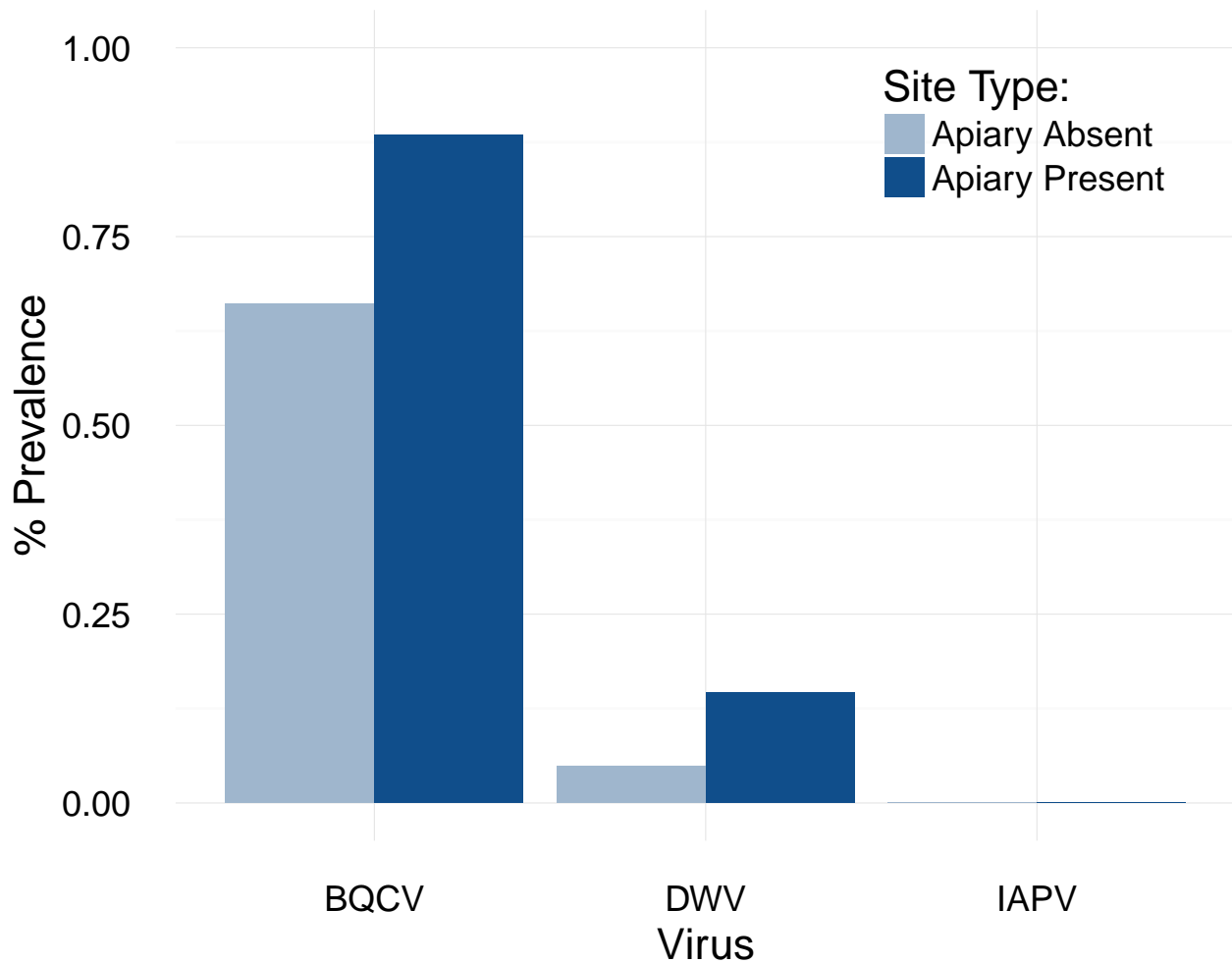
VirusSum1 <-VirusSum1[-c(7:9),]

#choosing color pallet
colors <- c("slategray3", "dodgerblue4")

#Create a bar graph for viruses by bombus species (aes= aesthetics):
plot1 <- ggplot(VirusSum1, aes(x=target_name, y=mean, fill=apiary_near_far)) +
  geom_bar(stat="identity",
    position=position_dodge()) + labs(x="Virus", y = "% Prevalence")

plot1 + theme_minimal(base_size = 17) + scale_fill_manual(values=colors, name="Site Type:", labels=c("A", "B"))

```



**Figure 2:** Percent prevalence of infected bumble bee individuals for black queen cell virus (BQCV), deformed wing virus (DWV) and Israeli acute paralysis virus (IAPV). Bumble bees were either caught in sites with honey bee apiaries present or no apiary nearby. BQCV ( $p < .0001$ ) and DWV ( $p = .00225$ ) were more prevalent in bumble bees caught in sites with a honey bee apiary present than in sites without an apiary nearby. IAPV was detected only in very low levels among a few individuals. This figure only includes bees with quantifiable viral infections above the threshold of detection for quantification.

```
#stats for % prevalence, apiary_near_far differences

#remove NAs before doing test
PresAbst_stat <- BombSurv[which(BombSurv$apiary_near_far!="NA"),]

PresAbst_stat <- split(PresAbst_stat, PresAbst_stat$target_name)

# chi.sq test for BQCV prev. vs apiary_near_far (p < 0.00001)
chisq.test(PresAbst_stat$BQCV$virusBINY, PresAbst_stat$BQCV$apiary_near_far)

##
## Pearson's Chi-squared test with Yates' continuity correction
##
## data: PresAbst_stat$BQCV$virusBINY and PresAbst_stat$BQCV$apiary_near_far
## X-squared = 19.225, df = 1, p-value = 1.162e-05
```

```

# chi.sq test for DWV prev. vs apiary_near_far (p = 0.00225)
chisq.test(PresAbst_stat$DWV$virusBINY, PresAbst_stat$DWV$apiary_near_far)

##
## Pearson's Chi-squared test with Yates' continuity correction
##
## data: PresAbst_stat$DWV$virusBINY and PresAbst_stat$DWV$apiary_near_far
## X-squared = 9.3334, df = 1, p-value = 0.00225

BombSurvSplit <- split(BombSurv, BombSurv$target_name)

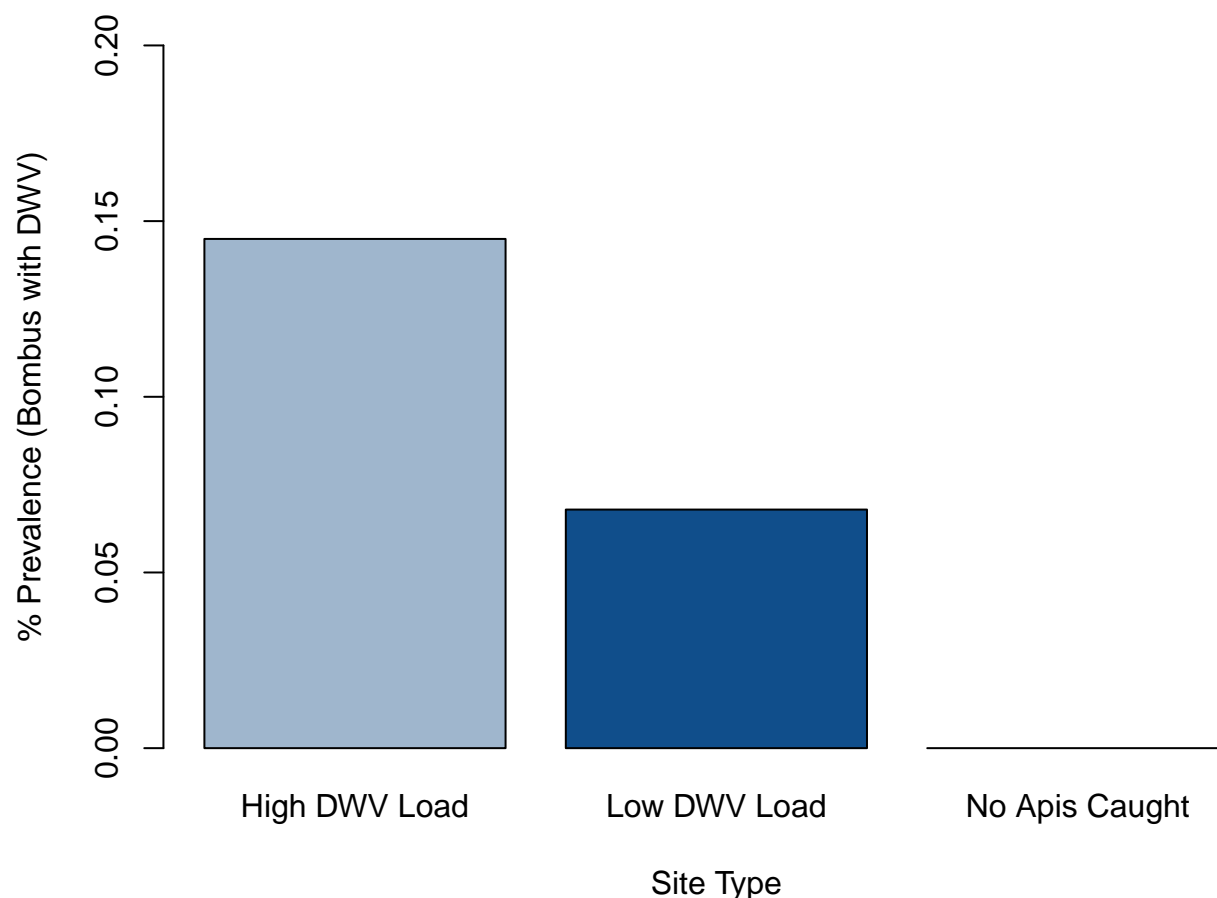
HBSiteSum <- ddply(BombSurvSplit$DWV, c("HBSiteBin", "target_name"), summarise,
  n = length(virusBINY),
  mean = mean(virusBINY, na.rm=TRUE),
  sd = sd(virusBINY, na.rm=TRUE),
  se = sd / sqrt(n))

colors <- c("slategray3", "dodgerblue4", "blue4")

barplot(height=HBSiteSum$mean,
  names.arg = c("High DWV Load",
    "Low DWV Load",
    "No Apis Caught"),
  xlab="Site Type",
  ylab="% Prevalence (Bombus with DWV)",
  ylim = c(0,0.2),
  main = "Bombus DWV Prevalence by Honey Bee DWV Load",
  col = colors
)

```

## Bombus DWV Prevalence by Honey Bee DWV Load



**Figure 3:** Percent prevalence for Bumble bees infected with deformed wing virus at sites where honey bees had high and low viral loads, and sites where no honey bees were present and therefore could not be collected. DWV was more prevalent in bumble bees caught at sites with honey bees with high average viral loads, than sites with honey bees with low average viral loads ( $p = 0.046$ ,  $X^2 = 3.9767$ ).

```
#remove NA rows from dataframe
BombSurvSplit <- BombSurvSplit$DWV[which(BombSurvSplit$DWV$HBSiteBin!="NA"),]
```

```
# chi.sq test for High and Low DWV load sites
chisq.test(BombSurvSplit$HBSiteBin, BombSurvSplit$virusBINY)
```

```
##
## Pearson's Chi-squared test with Yates' continuity correction
##
## data: BombSurvSplit$HBSiteBin and BombSurvSplit$virusBINY
## X-squared = 3.9767, df = 1, p-value = 0.04614
```

```
BombSurv$Date_collected <- as.Date(BombSurv$Date_collected, "%m/%d/%y")
```

```
tempsplit <- split(BombSurv, BombSurv$target_name)
```

```
BombSurv$Date_Pooled <- ifelse(BombSurv$Date_collected >= "2015-08-03", 90, ifelse(BombSurv$Date_collected < "2015-08-03", 0, 90))
```



```
library(dplyr)

sp <- split(BombSurv, BombSurv$species)

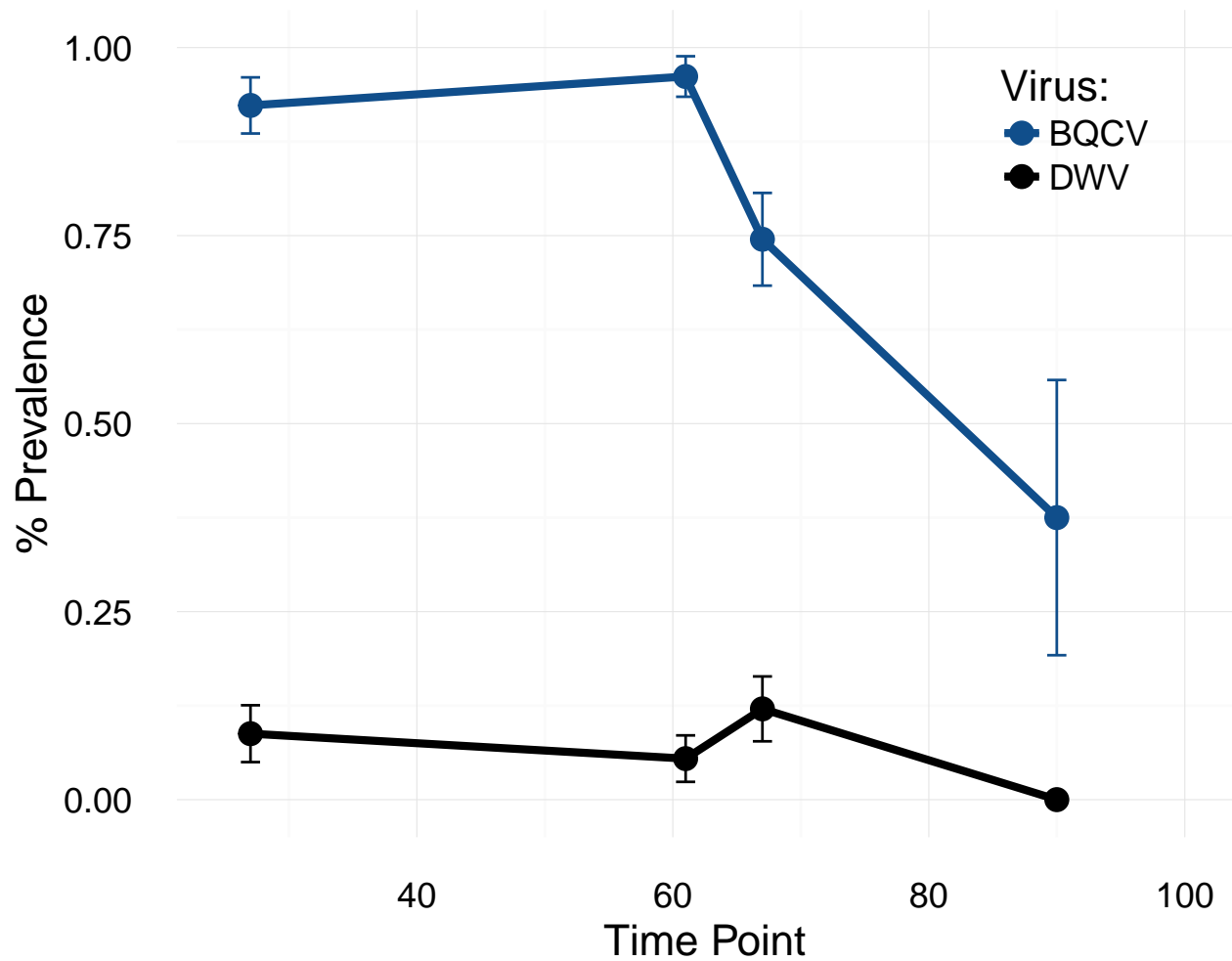
# create summary and sd and se using plyr
library(plyr)
VirusSum2 <- ddply(sp$Bimaculatus, c("target_name", "Date_Pooled"), summarise,
  n = length(virusBINY),
  mean = mean(virusBINY),
  sd = sd(virusBINY),
  se = sd / sqrt(n))

VirusSum2 <- VirusSum2[-(9:12),]

library(ggplot2)

#Create plot in ggplot
plot <- ggplot(data = VirusSum2,
  aes(x = Date_Pooled,
    y = mean,
    group = target_name,
    colour = target_name)
  ) + geom_line(size=1.5) + geom_point(size=4) + scale_colour_manual(values = c("dodgerblue", "red"))

# add a theme and add asterix for significance
plot + scale_fill_brewer(palette = "Paired") + theme_minimal(base_size = 17) + theme(legend.position=c("right", "top"))
```



```
#chisq.test(tempsplit$BQCV$Date_Pooled, tempsplit$BQCV$virusBINY)
```

```
mod <- glm(tempsplit$BQCV$virusBINY~tempsplit$BQCV$Date_collected)
```

```
num <- c(7, 3, 6, 4, 7, 3, 3, 5, 9, 11, 7, 6, 5, 4, 4)
sd(num)
```

```
## [1] 2.32379
```

Clustered HoneyBees:

```
head(BombSurv2)
```

```
##   X      site target_name sample_name Ct_mean Ct_sd quantity_mean
## 1 1 BOST-NEAR      DWV      283  32.615    NA          5.410
## 2 2 BOST-NEAR      DWV      210    NA     NA           NA
## 3 3 BOST-NEAR      DWV      299  32.604 0.159          5.463
## 4 4 BOST-NEAR      DWV      206    NA     NA           NA
## 5 5 BOST-NEAR      DWV      281  32.523    NA          5.742
## 6 6 BOST-NEAR      BQCV      204  28.339 0.038         75.704
##   quantity_sd run Date_collected      ID      species      host_plant
## 1           NA  57       6/18/15 BOSTN1 Bimaculatus Vicia cracca
```

```

## 2      NA 50      6/25/15 BOSTN25      Vagans Trifolium pratense
## 3      0.563 58      6/18/15 BOSTN15 Bimaculatus      Vicia cracca
## 4      NA 50      6/18/15 BOSTN20      Vagans      Vicia cracca
## 5      NA 57      6/18/15 BOSTN11 Bimaculatus      Vicia cracca
## 6      1.924 48      6/25/15 BOSTN30      Vagans Trifolium pratense
## date_processed pollen dil.factor apiary_near_far genome_copbee
## 1      4/18/16      1      10.0055      1      54129.75
## 2      2/22/16      1      8.7170      1      0.00
## 3      4/18/16      0      9.5460      1      52149.80
## 4      2/22/16      0      3.6350      1      0.00
## 5      4/18/16      1      9.0385      1      51899.07
## 6      2/22/16      1      6.7140      1      508276.66
## ACT_genome_copbee norm_genome_copbee norm_genome_copbeeHB Ct_mean_hb
## 1      1936335499      39283.02      1407310274      17.375
## 2      2456078210      0.00      1407310274      17.375
## 3      5279273065      13881.23      1407310274      17.375
## 4      79244690      0.00      1407310274      17.375
## 5      1987705795      36690.78      1407310274      17.375
## 6      895815651      797316.88      366994727      19.549
## HBCollected virusBINY HB_Abun HBSiteBin ShannonDIV SimpsonDIV
## 1      1      1 0.2857143      High      1.618767      0.7414096
## 2      1      0 0.2857143      High      1.618767      0.7414096
## 3      1      1 0.2857143      High      1.618767      0.7414096
## 4      1      0 0.2857143      High      1.618767      0.7414096
## 5      1      1 0.2857143      High      1.618767      0.7414096
## 6      1      1 0.2857143      <NA>      1.618767      0.7414096
## FisherDIV Density Migratory SBPV_PA SBPV_CPB ABPV_PA ABPV_CPB IAPV_PA
## 1      1.852017      40.8      0      0      0      0      0      0
## 2      1.852017      40.8      0      0      0      0      0      0
## 3      1.852017      40.8      0      0      0      0      0      0
## 4      1.852017      40.8      0      0      0      0      0      0
## 5      1.852017      40.8      0      0      0      0      0      0
## 6      1.852017      40.8      0      0      0      0      0      0
## IAPV_CPB DWV_PA DWV_CPB LSV.2_PA LSV.2_CPB CBPV_PA CBPV_CPB KBV_PA
## 1      0      1 2.85e+11      1 3.94e+09      1 415000      0
## 2      0      1 2.85e+11      1 3.94e+09      1 415000      0
## 3      0      1 2.85e+11      1 3.94e+09      1 415000      0
## 4      0      1 2.85e+11      1 3.94e+09      1 415000      0
## 5      0      1 2.85e+11      1 3.94e+09      1 415000      0
## 6      0      1 2.85e+11      1 3.94e+09      1 415000      0
## KBV_CPB VarroaTHR_PA Varroa100 NosemaPA NosemaCount Apiary_name
## 1      0      1      2.82      0      5      Bostwick
## 2      0      1      2.82      0      5      Bostwick
## 3      0      1      2.82      0      5      Bostwick
## 4      0      1      2.82      0      5      Bostwick
## 5      0      1      2.82      0      5      Bostwick
## 6      0      1      2.82      0      5      Bostwick

```

```

CopDist <- ddply(BombSurv2, c("target_name", "site"), summarise,
  n = length(norm_genome_copbeeHB),
  mean = mean(norm_genome_copbeeHB, na.rm=TRUE),
  sd = sd(norm_genome_copbeeHB, na.rm=TRUE),
  se = sd / sqrt(n))

```

```
CopDist <- CopDist[c(23:44),]
```

```
hist(log(CopDist$mean),  
     breaks=20,  
     xlim = c(5,25),  
     col = "dodgerblue4",  
     xlab = "Honeybee DWV level log(viral load)",  
     main = "Distribution of DWV Viral Load in Honeybees",  
     cex.lab = 1.3,  
     font.lab = 2  
     )
```

