# Athens Dengue Data

Michelle Evans

April 17, 2017

This document goes over the main results from the 2016 Athens Dengue Field Project.

Infection Dynamics

Methods Overview: Mosquitoes were reared at 9 sites (3 rural, 3 surburban, 3 urban) and brought back to the lab to infect with dengue. Mosquitoes were reared in four trays at each site, but pooled by site for infections because there were not enough numbers. At 21 days post infection, they were processed to test for infection (body), dissemination (head), and infectiousness (saliva). This was repeated in the fall and summer. Climate data reported was averaged at the site level over the larval rearing period, and weighted to account for the amount of time mosquitoes spent in each tray.

Infection by Class and Season

The first analysis was to determine how infection dynamics differed by class (rural, suburban, or urban) & season (summer or fall).
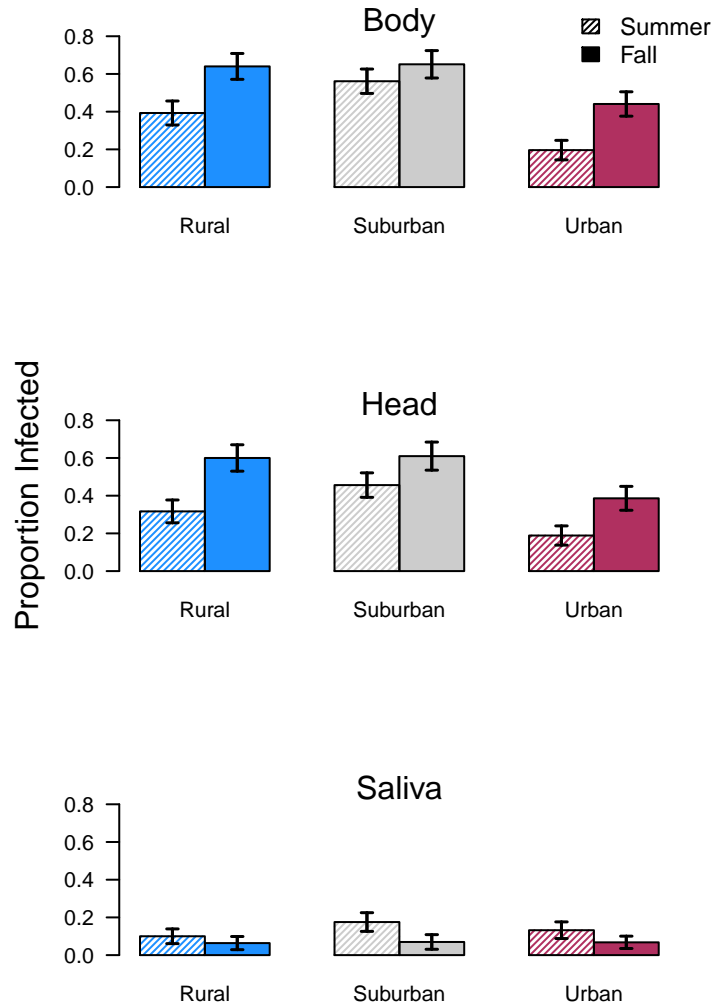
A binomial generalized linear mixed-effects model was fit to the data, with class and season and their interaction as fixed effects and site as a random effect. To test for significance, we calculated an analysis of variance table using a Type II Wald chisquare test for the fixed effects. Both class and season had significant effects on body and head infections, however there was no effect of either on saliva infection.

We found that infection in the body and head was significantly lower in the summer than the fall. Although there was a trend towards higher infection in the saliva in the summer, this was not significant (p=0.05676). Land class also influence infection of both the body and head, with urban mosquitoes having lower rates of infection than rural and suburban mosquitoes. Saliva infections were unaffected by land class, however.

	Chisq	Df	Pr(>Chisq)
class	18.73327	2	8.55e-05
block	12.60856	1	3.84e-04

As both the summer season and urban land classes had the highest temperatures in their respecitive groups, this suggests that these differences in vector competence may be due to differences in temperature in the larval environment.

Infection and Microclimate

```
## [1] "2016-08-01 17:04:48 EDT" "2016-08-05 09:14:48 EDT"
```

```
## [1] "2016-08-01 17:07:27 EDT" "2016-11-07 16:51:04 EST"
```

Based on our findings above, we wanted to investigate if these patterns across season and class could be driven by microclimate, and how.

Becuase our experimental design was somewhat limited by the number of mosquitoes emerging, we grouped infections by site, rather than tray. This means that the microclimate data used is the same for all mosquitoes across sites, and that we cannot use it to predict individual level infections in a meaningful way, as we did for land class and season. Therefore, our response variable in this analysis is the proportion of infected mosquitoes by site. This greatly reduces our sample size to 18, and therefore our statistical power.

The original starting set of variables were the average daily mean, minimum, and maximum temperature and relative humidity per site, daily temperature range, and the average number of hours below 16C and above 31C. These temperature values were chosen based on thresholds from the Mordecai et al 2017 model of temperature dependence of dengue infections. Because maximum relative humidity is 100% for all sites, this variable was dropped. This resulted in eight total environmental covariates. Becuase temperature and relative humidity are so dependent on each other, this resulted in high correlations amongst the covariates.

To select covariates, I used a lasso regression in the `glmnet` package to conduct a stepwise regression selection process based on the RMSE. At each step, the variable whose omission resulted in the greatest drop in RMSE was dropped from the model. As the selection process continued, previously dropped variables were again included in the model at each step to determine any effect on RMSE, and if this increased the error, were left out of the final model. Variable selection ended when the dropping of any of the included variables did not decrease the RMSE of the overall model.

Because most of these variables were still highly correlated ($\rho > 0.9$), I then further selected models to eliminate collinearity amongst covariates. For this stage, I chose to use a linear mixed model because many of the exploratory visualizations were clearly linear, but with a strong random effect of season, which I could include in a mixed model. I modeled all variables individually and with an interaction term between another, if the correlation between the variables was less than 0.75. Models were chosen based on that which minimized the AICc and maximized the marginal r^2.

Body Infection

LASSO regression identified mean, minimum, and maximum temperatures and daily temperature range as important variables, with mean temperature the most important.

The final linear mixed model included mean temperature as a fixed effect. It had a marginal r^2 of 0.442. It also had the lowest AIC value out of all models.

The relationship between mean temperature and body infection is negative (β=-0.086) and significant ($p < 0.05$). Higher temperatures seem to be leading to lower body infection rates. Additionally, the relationship seems qualitatively different over the seasons, with the season as a random intercept explaining nearly as much of the variance as the mean temperature ($r^2_{GLMM_c}$=0.885. This may be because the summer temperatures are near the thermal optimum, where the effect of temperature is more pronounced.
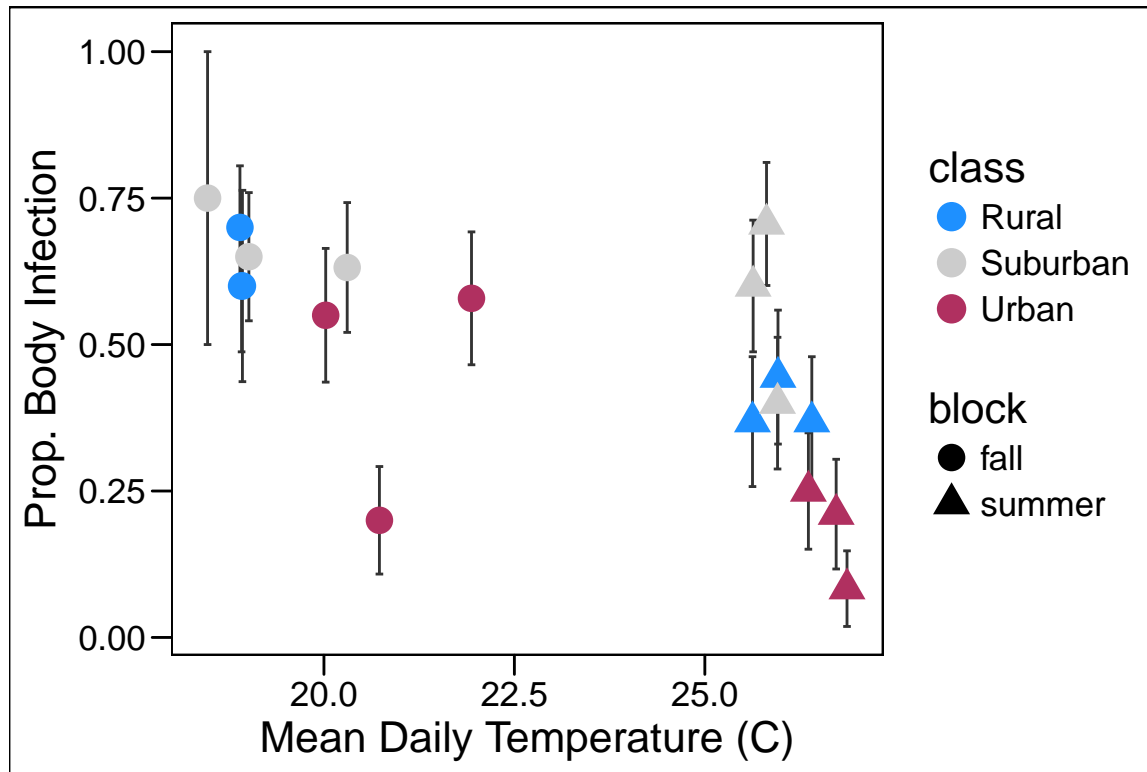
Head Infection

The final model included maximum and mean temperature and mean relative humidity. The most important variable was maximum temperature (Figure 4), however the difference in variable importance was less than in the Body regression. Although the mean relative humidity is included in this model, it is relatively unimportant. This leads me to conclude that temperature is also the main driver in this model, as it was for body infection.

As above, we found that head infection rates decreased with increasing temperatures, particularly the maximum temperature (Figure 5). Again, the relationship differs slightly across the seasons.

Saliva Infection

While saliva had no significant differences across land class, there was a slight trend towards higher saliva in the summer, although not significant. Therefore, I thought it would be informative to also look at the effect of microclimate on saliva infection. Using similar methods above, the final model included maximum temperature, mean temperature, and mean relative humidity (Figure 6). Interestingly the relationship between temperature and saliva infection is contrary to that seen for body and head infection, that is, it is positive. At higher temperatures, saliva infection rates were higher, although the effect is not as large as those seen earlier (Figure 7).
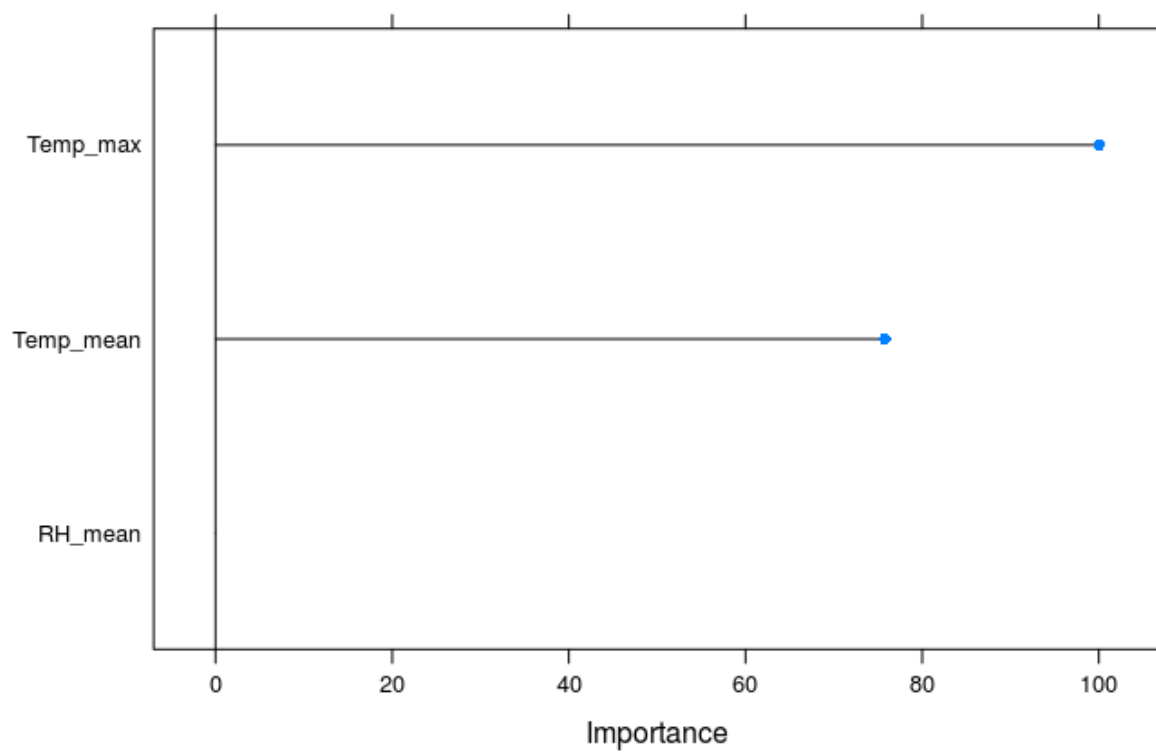
Figure 1: Variable importance plot from final head glmnet model. Importance values are scaled.
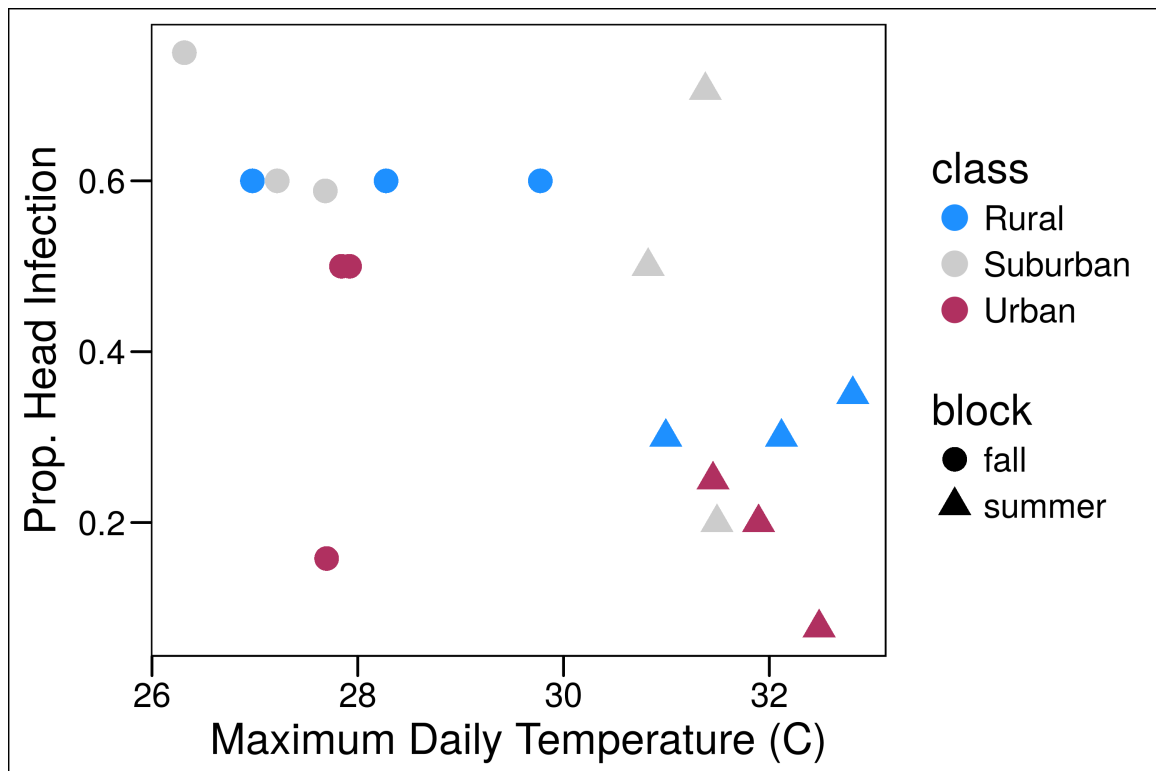
Figure 2: Head infection across maximum daily temperatures in the summer (triangle) and fall (circle) replicates. Colors represent land class.
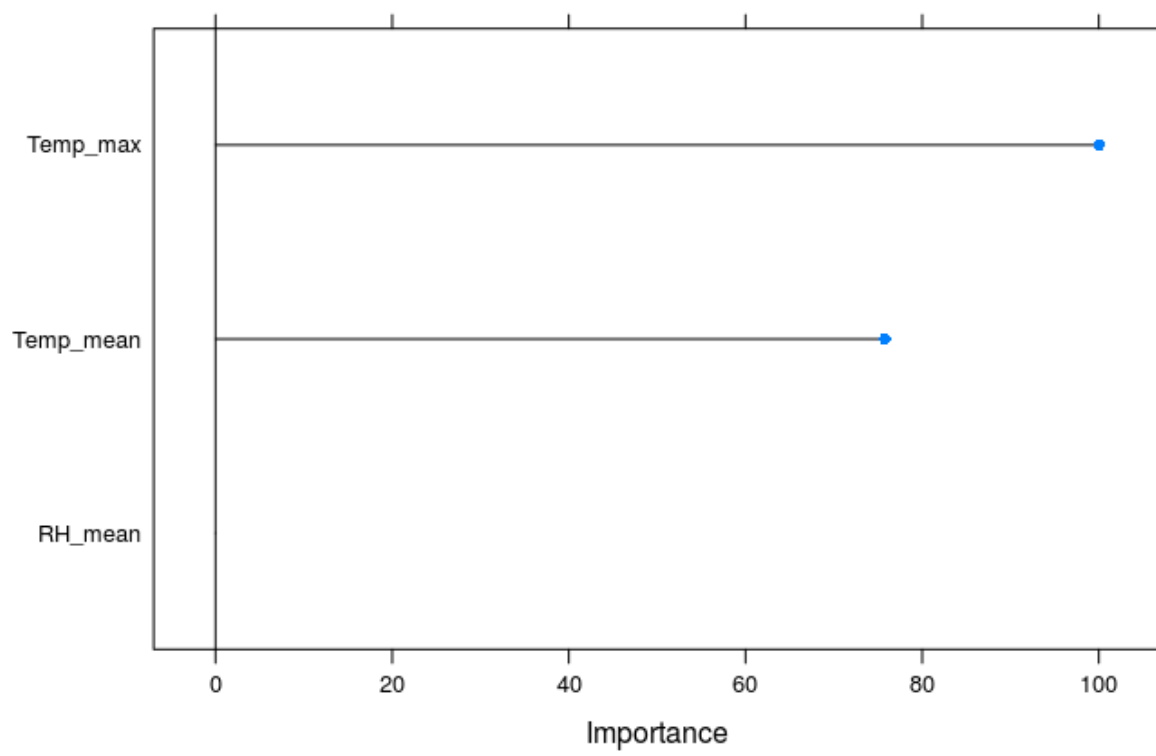
Figure 3: Variable importance plot from final head glmnet model. Importance values are scaled.
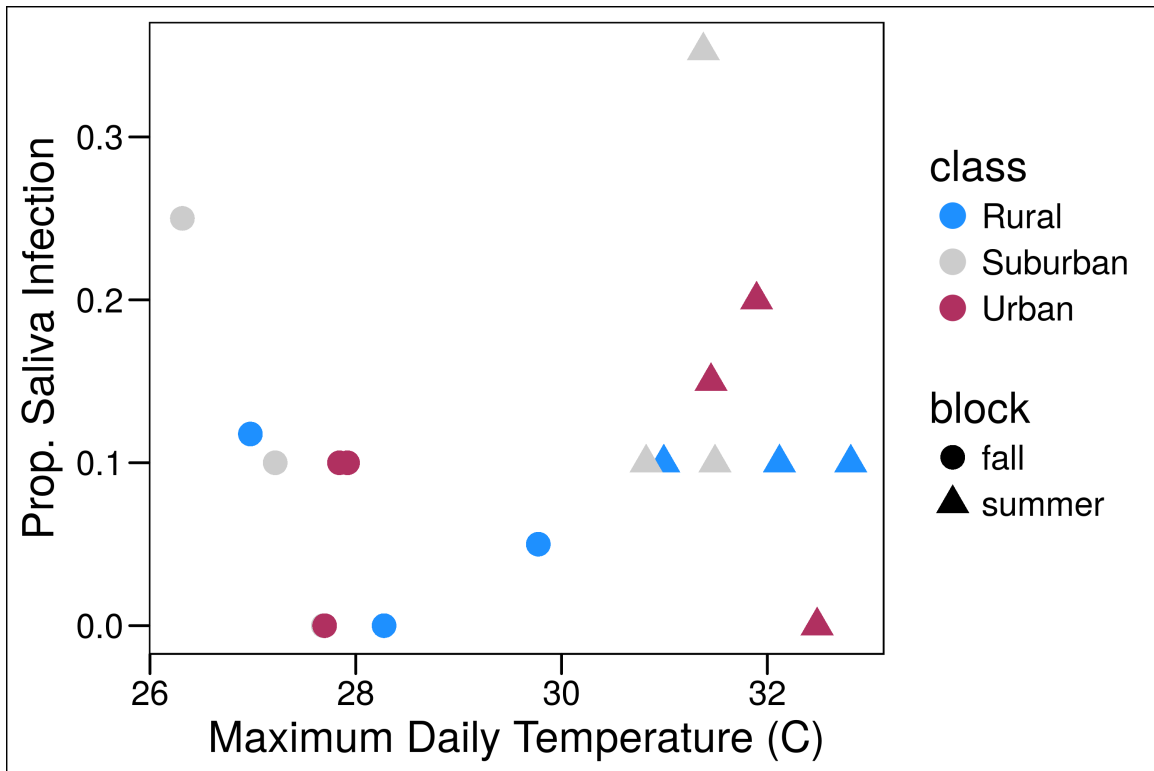
Figure 4: Saliva infection across maximum daily temperatures in the summer (triangle) and fall (circle) replicates. Colors represent land class.