Lab 9

DO NOT INCLUDE NAMES - Just add names in gradescope

# Rules

In groups of 2 or 3, complete the following.

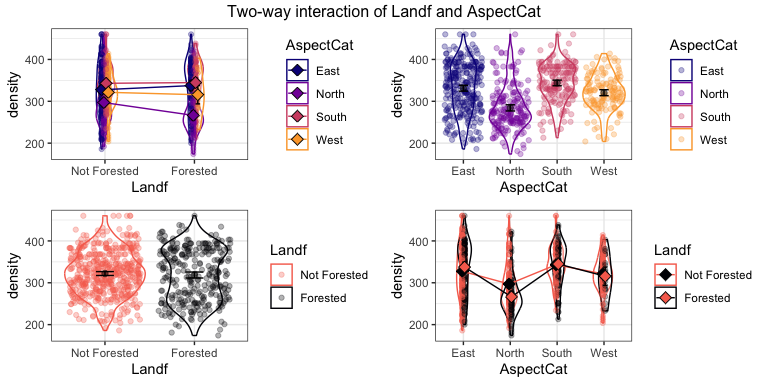
# Modeling Snow Density

We will be using the data set from Wetlaufer, Hendrikx, and Marshall (2016) - study where they explored the relationship between snow density () or snow depth (snow, mm) or snow water equivalent (SWE, mm) with a suite of predictor variables. In this lab we will condition on there being snow present and then try to model snow density with models that will compete with what they discuss in Tables 3 and 4. We will be interested in using elev (Elevation, m), Land (forest cover with 0 = unforested and 10 = forested), rad (Potential Solar radiation, ), curvature (see <https://blogs.esri.com/esri/arcgis/2010/10/27/understanding-curvature-rasters/> for a description), aspect (orientation of slope in degrees (0 to 360 degrees)), and angle (angle of slope in degrees with 0 being flat) as predictors. Also pay attention to the strata variable (read its definition in the paper) and the role that played in the data collection and should in the analysis, as we will revisit this.

* Wetlaufer, K., Hendrikx, J., and L. Marshall (2016) Spatial Heterogeneity of Snow Density and Its Influence on Snow Water Equivalence Estimates in a Large Mountainous Basin. *Hydrology*, 3(1):3, <doi:10.3390/hydrology3010003>. Available at <http://www.mdpi.com/2306-5338/3/1/3/htm> and on D2L

Run the following code to get re-started with the data set.

data(snowdepths)  
snowdepths <- snowdepths %>%  
 mutate(AspectCat = factor(case\_when(  
 aspect %in% (0:45)~ "North",  
 aspect %in% (315:360)~ "North",  
 aspect %in% 45:(90+45) ~ "East",  
 aspect %in% (90+45):(180+45) ~ "South",  
 aspect %in% (180+45):315 ~ "West"  
 )),  
 SnowPresence = factor(case\_when(  
 snow == 0 ~ "None",  
 snow > 0 ~ "Some"  
 )),  
 Landf = factor(cover),  
 Landf = fct\_recode(Landf,  
 "Not Forested" = "0",  
 "Forested" = "10")  
 )  
  
snowdepthsR <- snowdepths %>% drop\_na(density)  
  
ggintplot(response = "density", groupvars = c("Landf", "AspectCat"), data = snowdepthsR)



lmA <- lm(density ~ Landf\*AspectCat, data = snowdepthsR)  
Anova(lmA)

## Anova Table (Type II tests)  
##   
## Response: density  
## Sum Sq Df F value Pr(>F)  
## Landf 3875 1 1.4629 0.2268340  
## AspectCat 388163 3 48.8448 < 2.2e-16  
## Landf:AspectCat 47814 3 6.0167 0.0004707  
## Residuals 2079429 785

lmA %>% tbl\_regression(intercept = T)

| **Characteristic** | **Beta** | **95% CI***1* | **p-value** |
| --- | --- | --- | --- |
| (Intercept) | 328 | 320, 335 | <0.001 |
| Landf |  |  |  |
| Not Forested | — | — |  |
| Forested | 9.9 | -2.6, 22 | 0.12 |
| AspectCat |  |  |  |
| East | — | — |  |
| North | -31 | -43, -18 | <0.001 |
| South | 15 | 2.5, 28 | 0.020 |
| West | -6.1 | -18, 6.0 | 0.3 |
| Landf \* AspectCat |  |  |  |
| Forested \* North | -41 | -60, -21 | <0.001 |
| Forested \* South | -8.1 | -28, 11 | 0.4 |
| Forested \* West | -16 | -43, 11 | 0.3 |
| *1*CI = Confidence Interval | | | |

The estimated model in lmA is

1. Use the provided model to estimate the mean for a nonforested site that has an East aspect. Show/explain your work.

**The estimated mean for a nonforested site with an East aspect is 328. Nonforested and East are both baselines for the categorical variables, so there are no indicator variables associated with them. Thus, we are left with the intercept.**

328

## [1] 328

1. Use the provided model to estimate the mean for a nonforested site that has a North aspect. Show/explain your work.

**The estimated mean for a nonforested site with an East aspect is 297. Nonforested is the baseline for the categorical variable, so the indicator for forested is set to 0. Thus, we are left with the intercept and the beta value associated with the indicator for North aspect.**

328 - 31

## [1] 297

1. The estimated slope for of the North indicator variable is -30.5. Write a size interpretation for that difference, using your previous work to help with being specific about what the slope estimates and remembering to add a 95% CI for that slope coefficient. Also note that the units for snow density are .

**\*\* For two otherwise similar sites that are both nonforested but that differ in Aspect, the estimated mean density for the North aspect sites is 30.5 kg/m^3 lower than for the East site (95% CI of 18 to 43).**

**For two otherwise similar sites that are nonforested but differ in Aspect, the estimated mean snow density for the North aspect site is 30.5 lower than the East aspect site. (95% CI: -43,-18)**

1. The following code generates Tukey’s HSD results for the interaction model. Find the same estimated difference that you reported in the previous question (sign might be different) and report how that result is labeled in the provided output. Then compare the provided Tukey’s familywise adjusted confidence interval to the previous result. How can you explain the difference?

**The difference between nonforestedEast - nonforestedNorth in the Tukey is 30.52. The sign difference is because they were taking East - North, whereas above it was North - East. The previously reported confidence interval, translated into East - North, was 18 to 43. The tukey reported confidence interval was from 11.719, 49.33. This confidence interval was wider because Tukey adjusts for multiple comparisons. The previous confidence interval was narrower because the error was not adjusted for these multiple comparisons.**

emA <- emmeans(lmA, pairwise ~ Landf \* AspectCat, adjust = "tukey")  
summary(emA)

## $emmeans  
## Landf AspectCat emmean SE df lower.CL upper.CL  
## Not Forested East 328 3.88 785 320 335  
## Forested East 337 5.05 785 328 347  
## Not Forested North 297 4.82 785 288 307  
## Forested North 266 5.72 785 255 278  
## Not Forested South 343 5.34 785 333 353  
## Forested South 345 5.49 785 334 356  
## Not Forested West 321 4.78 785 312 331  
## Forested West 316 11.20 785 294 338  
##   
## Confidence level used: 0.95   
##   
## $contrasts  
## contrast estimate SE df t.ratio p.value  
## Not Forested East - Forested East -9.92 6.37 785 -1.559 0.7748  
## Not Forested East - Not Forested North 30.52 6.19 785 4.933 <.0001  
## Not Forested East - Forested North 61.19 6.91 785 8.854 <.0001  
## Not Forested East - Not Forested South -15.43 6.60 785 -2.339 0.2738  
## Not Forested East - Forested South -17.22 6.72 785 -2.562 0.1716  
## Not Forested East - Not Forested West 6.09 6.16 785 0.989 0.9760  
## Not Forested East - Forested West 11.85 11.90 785 0.998 0.9748  
## Forested East - Not Forested North 40.45 6.98 785 5.795 <.0001  
## Forested East - Forested North 71.11 7.63 785 9.323 <.0001  
## Forested East - Not Forested South -5.51 7.35 785 -0.750 0.9954  
## Forested East - Forested South -7.29 7.45 785 -0.978 0.9774  
## Forested East - Not Forested West 16.01 6.95 785 2.303 0.2931  
## Forested East - Forested West 21.78 12.30 785 1.769 0.6417  
## Not Forested North - Forested North 30.66 7.48 785 4.099 0.0012  
## Not Forested North - Not Forested South -45.96 7.19 785 -6.390 <.0001  
## Not Forested North - Forested South -47.74 7.30 785 -6.537 <.0001  
## Not Forested North - Not Forested West -24.44 6.79 785 -3.600 0.0081  
## Not Forested North - Forested West -18.67 12.20 785 -1.528 0.7924  
## Forested North - Not Forested South -76.62 7.82 785 -9.795 <.0001  
## Forested North - Forested South -78.40 7.92 785 -9.893 <.0001  
## Forested North - Not Forested West -55.10 7.45 785 -7.394 <.0001  
## Forested North - Forested West -49.33 12.60 785 -3.914 0.0025  
## Not Forested South - Forested South -1.78 7.65 785 -0.233 1.0000  
## Not Forested South - Not Forested West 21.52 7.16 785 3.004 0.0554  
## Not Forested South - Forested West 27.29 12.40 785 2.194 0.3562  
## Forested South - Not Forested West 23.30 7.28 785 3.203 0.0306  
## Forested South - Forested West 29.07 12.50 785 2.326 0.2809  
## Not Forested West - Forested West 5.77 12.20 785 0.473 0.9998  
##   
## P value adjustment: tukey method for comparing a family of 8 estimates

confint(emA)

## $emmeans  
## Landf AspectCat emmean SE df lower.CL upper.CL  
## Not Forested East 328 3.88 785 320 335  
## Forested East 337 5.05 785 328 347  
## Not Forested North 297 4.82 785 288 307  
## Forested North 266 5.72 785 255 278  
## Not Forested South 343 5.34 785 333 353  
## Forested South 345 5.49 785 334 356  
## Not Forested West 321 4.78 785 312 331  
## Forested West 316 11.20 785 294 338  
##   
## Confidence level used: 0.95   
##   
## $contrasts  
## contrast estimate SE df lower.CL upper.CL  
## Not Forested East - Forested East -9.92 6.37 785 -29.268 9.42  
## Not Forested East - Not Forested North 30.52 6.19 785 11.719 49.33  
## Not Forested East - Forested North 61.19 6.91 785 40.184 82.19  
## Not Forested East - Not Forested South -15.43 6.60 785 -35.484 4.62  
## Not Forested East - Forested South -17.22 6.72 785 -37.637 3.21  
## Not Forested East - Not Forested West 6.09 6.16 785 -12.621 24.79  
## Not Forested East - Forested West 11.85 11.90 785 -24.258 47.97  
## Forested East - Not Forested North 40.45 6.98 785 19.236 61.66  
## Forested East - Forested North 71.11 7.63 785 47.928 94.29  
## Forested East - Not Forested South -5.51 7.35 785 -27.833 16.81  
## Forested East - Forested South -7.29 7.45 785 -29.949 15.36  
## Forested East - Not Forested West 16.01 6.95 785 -5.115 37.13  
## Forested East - Forested West 21.78 12.30 785 -15.645 59.20  
## Not Forested North - Forested North 30.66 7.48 785 7.931 53.39  
## Not Forested North - Not Forested South -45.96 7.19 785 -67.812 -24.10  
## Not Forested North - Forested South -47.74 7.30 785 -69.936 -25.54  
## Not Forested North - Not Forested West -24.44 6.79 785 -45.067 -3.81  
## Not Forested North - Forested West -18.67 12.20 785 -55.814 18.47  
## Forested North - Not Forested South -76.62 7.82 785 -100.390 -52.84  
## Forested North - Forested South -78.40 7.92 785 -102.486 -54.32  
## Forested North - Not Forested West -55.10 7.45 785 -77.749 -32.45  
## Forested North - Forested West -49.33 12.60 785 -87.634 -11.03  
## Not Forested South - Forested South -1.78 7.65 785 -25.046 21.48  
## Not Forested South - Not Forested West 21.52 7.16 785 -0.254 43.29  
## Not Forested South - Forested West 27.29 12.40 785 -10.505 65.08  
## Forested South - Not Forested West 23.30 7.28 785 1.189 45.41  
## Forested South - Forested West 29.07 12.50 785 -8.918 67.06  
## Not Forested West - Forested West 5.77 12.20 785 -31.326 42.86  
##   
## Confidence level used: 0.95   
## Conf-level adjustment: tukey method for comparing a family of 8 estimates

1. Run pwpm on the emmeans version of the model with the Tukey’s adjusted results. Use it to find the difference in the estimated mean for Forested West and Not Forested West and the Tukey’s adjusted p-value for that difference. No other discussion, just use it to find those results.

**Difference in estimated mean: 5.77** **for not forested west – forested west. P-value = 0.9988.**

pwpm(emA)

## Not Forested East Forested East Not Forested North  
## Not Forested East [328] 0.7748 <.0001  
## Forested East -9.92 [337] <.0001  
## Not Forested North 30.52 40.45 [297]  
## Forested North 61.19 71.11 30.66  
## Not Forested South -15.43 -5.51 -45.96  
## Forested South -17.22 -7.29 -47.74  
## Not Forested West 6.09 16.01 -24.44  
## Forested West 11.85 21.78 -18.67  
## Forested North Not Forested South Forested South  
## Not Forested East <.0001 0.2738 0.1716  
## Forested East <.0001 0.9954 0.9774  
## Not Forested North 0.0012 <.0001 <.0001  
## Forested North [266] <.0001 <.0001  
## Not Forested South -76.62 [343] 1.0000  
## Forested South -78.40 -1.78 [345]  
## Not Forested West -55.10 21.52 23.30  
## Forested West -49.33 27.29 29.07  
## Not Forested West Forested West  
## Not Forested East 0.9760 0.9748  
## Forested East 0.2931 0.6417  
## Not Forested North 0.0081 0.7924  
## Forested North <.0001 0.0025  
## Not Forested South 0.0554 0.3562  
## Forested South 0.0306 0.2809  
## Not Forested West [321] 0.9998  
## Forested West 5.77 [316]  
##   
## Row and column labels: Landf:AspectCat  
## Upper triangle: P values adjust = "tukey"  
## Diagonal: [Estimates] (emmean)   
## Lower triangle: Comparisons (estimate) earlier vs. later

1. The evidence drops for the potential interaction if we account for elevation and radiation (results provided below). We likely would drop the interaction in the following model. The additive version of the model is also provided and Tukey’s HSD is provided for the AspectCat variable. Report a size interpretation for the North versus West aspects from the additive model (lmB).

**North - West = -12.662 (-26.74, 1.41)**

**For two otherwise similar sites but that differ on Aspect, the estimated mean density for a West site is 12.662 higher than for a North site (95% Tukey’s adjusted CI of -1.41 to 26.74), controlled for land cover, radiation, and elevation.**

**For two otherwise similar sites but differ in aspect, the estimated mean snow density for the North aspect is 12.662 less than a West aspect. (95% CI: -26.74, 1.41)**

lmA <- lm(density ~ Landf\*AspectCat + rad + elev, data = snowdepthsR)  
Anova(lmA)

## Anova Table (Type II tests)  
##   
## Response: density  
## Sum Sq Df F value Pr(>F)  
## Landf 1073 1 0.5559 0.4561  
## AspectCat 11715 3 2.0240 0.1091  
## rad 97693 1 50.6355 2.515e-12  
## elev 533090 1 276.3085 < 2.2e-16  
## Landf:AspectCat 9615 3 1.6613 0.1739  
## Residuals 1510665 783

lmB <- lm(density ~ Landf + AspectCat + rad + elev, data = snowdepthsR)  
Anova(lmB)

## Anova Table (Type II tests)  
##   
## Response: density  
## Sum Sq Df F value Pr(>F)  
## Landf 1073 1 0.5545 0.4567  
## AspectCat 11715 3 2.0189 0.1098  
## rad 109560 1 56.6434 1.432e-13  
## elev 560398 1 289.7315 < 2.2e-16  
## Residuals 1520280 786

emB <- emmeans(lmB, pairwise ~ AspectCat, adjust = "tukey")  
confint(emB)

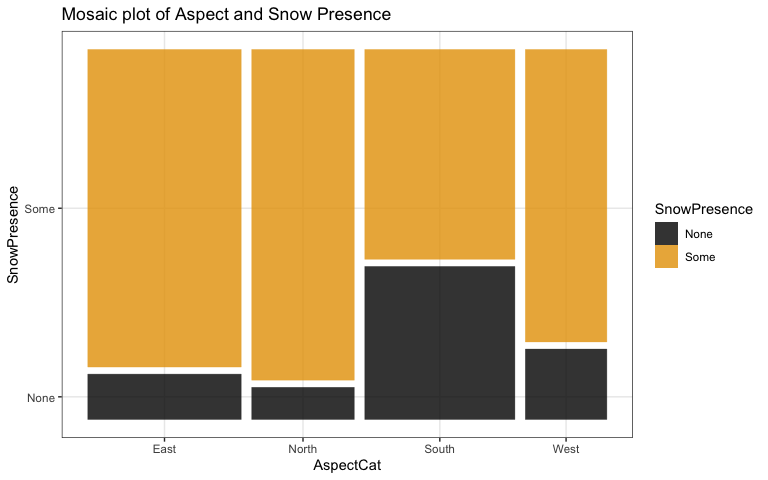
## $emmeans  
## AspectCat emmean SE df lower.CL upper.CL  
## East 322 2.71 786 317 327  
## North 312 3.99 786 304 320  
## South 324 3.91 786 316 332  
## West 325 3.94 786 317 332  
##   
## Results are averaged over the levels of: Landf   
## Confidence level used: 0.95   
##   
## $contrasts  
## contrast estimate SE df lower.CL upper.CL  
## East - North 9.821 4.92 786 -2.85 22.49  
## East - South -2.215 4.67 786 -14.25 9.82  
## East - West -2.841 4.72 786 -14.98 9.30  
## North - South -12.036 6.44 786 -28.61 4.53  
## North - West -12.662 5.47 786 -26.74 1.41  
## South - West -0.626 5.60 786 -15.03 13.78  
##   
## Results are averaged over the levels of: Landf   
## Confidence level used: 0.95   
## Conf-level adjustment: tukey method for comparing a family of 4 estimates

## Logistic regression

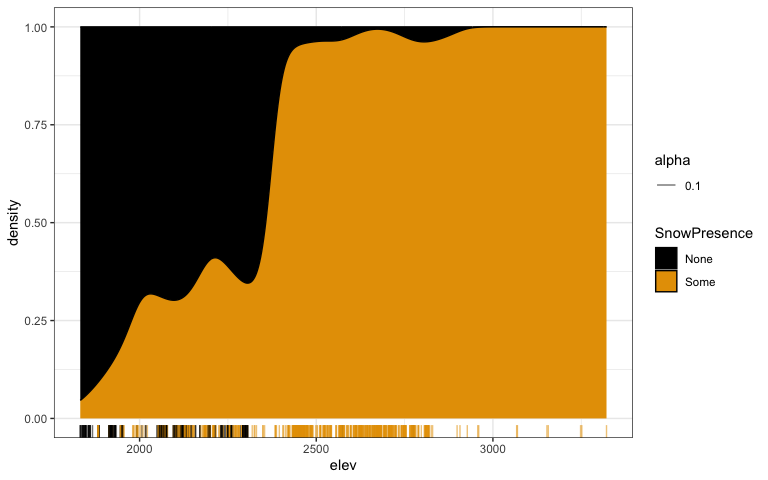
In the previous work, the missing values on snow density were removed because the locations did not have snow to measure. Let’s consider the presence of snow as a response variable, something that might be of interest for a skier at this time of the year… To work with this, we need to go back to the full data set, snowdepths.

1. Uncomment the provided templates for making a mosaic plot and a conditional density plots. Use the plots to make the appropriate display of the presence of snow both as a function of elev and AspectCat (two plots). Discuss the plot for the elevation predictor and whether this is what you might expect?

snowdepths %>%   
 ggplot() +   
 geom\_mosaic(aes(x = product(SnowPresence, AspectCat), fill = SnowPresence), offset = 0.02) +   
 scale\_fill\_colorblind() +   
 labs(title = "Mosaic plot of Aspect and Snow Presence")



snowdepths %>% ggplot(aes(x = elev, fill = SnowPresence)) +   
 geom\_density(position='fill', bw =50) + #Add bw = option to avoid abstract art  
 scale\_fill\_colorblind() +   
 geom\_rug(aes(col = SnowPresence, alpha = 0.1)) +  
 scale\_color\_colorblind()



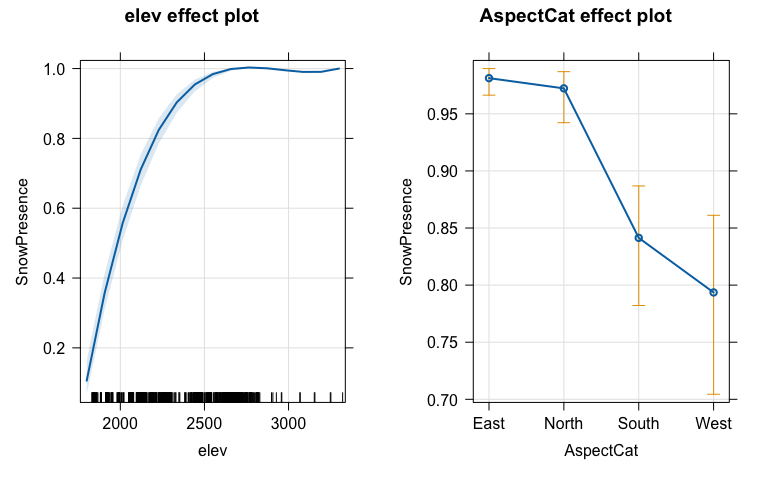
#Pick bandwidth (bw) to look "nice" (no white patches) but not smooth too much

1. Based on the provided results, complete writing out the estimated model, including using the results from the model and the previous plot(s) to define what a success is in the model:

glm1 <- glm(SnowPresence ~ elev + AspectCat, data = snowdepths, family = binomial)  
glm1 %>% tbl\_regression(intercept = T, exponentiate = F)

| **Characteristic** | **log(OR)***1* | **95% CI***1* | **p-value** |
| --- | --- | --- | --- |
| (Intercept) | -17 | -20, -14 | <0.001 |
| elev | 0.01 | 0.01, 0.01 | <0.001 |
| AspectCat |  |  |  |
| East | — | — |  |
| North | -0.40 | -1.1, 0.33 | 0.3 |
| South | -2.3 | -2.8, -1.8 | <0.001 |
| West | -2.6 | -3.4, -1.9 | <0.001 |
| *1*OR = Odds Ratio, CI = Confidence Interval | | | |

plot(allEffects(glm1), grid = T, type = "response")



* where snow presence is a success.

1. Report any resources outside your group and the course provided materials that you used and how they impacted your answers. Report NONE if there were none.

**NONE**