Welcome to The Carpentries Etherpad!

This pad is synchronized as you type, so that everyone viewing this page sees the same text. This allows you to collaborate seamlessly on documents.

Use of this service is restricted to members of The Carpentries community; this is not for general purpose use (for that, try https://etherpad.wikimedia.org).

Users are expected to follow our code of conduct: https://docs.carpentries.org/topic_folders/policies/code-of-conduct.html



---

Please tell us who you are?
Jon Jablonski | l Librarian | B&J half baked brownie
Echelle Burns || Project Researcher at Environmental Markets Lab (emLab)Ac at UC Santa Barbara || coffee oreo ice cream!
Gavin McDonald || Project Researcher at Environmental Markets Lab at UCSB || salted caramel
Amelia Ritger | EEMB Graduate Student | Chocolate peanut butter cup
Julia || MEDS (environmental data science) Grad student || I don't eat ice cream, nooooooo.
Daphne || NCEAS staff (Bren Alum, what's up brennies) || daiquiri ice cream from Baskin Robbins (this is a great flavor)
Karina || PhD Student and Science Director for an NGO || cookies + cream ice cream!
Kristi Liu | Library | really craving one of those samanco fish icecreams right now
Marie | environmental data science student | peanut butter cup
Amelia Meyer || Statistics and Data Science Student || cotton candy or peanut butter
Wayne Johnson - Third Year Transfer student AA Psychology - SEEDS program - Cookie Dough
Camila Vargas - Research Assistant MSI, UCSB - Pistaccio
Jenny Couture, fishery analyst, coffee oreo
Xiangmin (Sam) Sun || postdoc at Ilinois ||
Mansi (undergrad in envs & geog) doiry-free chocolate brownie
Chris Honeyman || Project Scientist at MSI || Chocolate Chip Cookie Dough
Matteen Atiqi || 3rd-year Stats & DS / SEEDS program || cookies and cream
Pike Spector || Marine scientist (NOAA) || kitchen sink ice cream! e.g. Ben and Jerries ;-)
Taylor Medina || MESM student || coffee
Arnav Kumar | Computer Science | Mint
Nick McManus || MESM Graduate student || PB cup
Ava Ge || Financial Math & Stats Student || Chocolate and straberry
Christine Afework| Third Year Undergrad Psychology (SEEDS)| cookies and cream
Jason Flower | Researcher at emlab (Environmental Markets Lab) at UCSB | chocolate
Merhawi GebreMichael | Postdoc| chocolate
Hally Zhou | Econ & GIS Undergrad | Matcha
Kaili Brande - PhD student at Bren School - cookies and cream ice cream
Johanna || PhD candidate EEMB || chocolate
Althea Marks | Data Fellow NCEAS | Mint chip

##########

### Glossary

GIS
raster: a data file that is a rectangular array of pixels
pixels = picture element. One dot in an image.
cells = pixels
crs = coordinate reference system = projection ~= proj4 file

### External Resources

NCEAS color cheatsheet
https://www.nceas.ucsb.edu/sites/default/files/2020-04/colorPaletteCheatsheet.pdf

Another cheat sheet:
http://sape.inf.usi.ch/quick-reference/ggplot2/colour

about virdis:
https://cran.r-project.org/web/packages/viridis/vignettes/intro-to-viridis.html)

Raster data:
https://desktop.arcgis.com/en/arcmap/latest/manage-data/raster-and-images/what-is-raster-data.htm

links to full driver documentation:
https://gdal.org/drivers/vector/index.html

overview of geospatial definitions/concepts from NCEAS:
https://www.nceas.ucsb.edu/sites/default/files/2020-04/OverviewCoordinateReferenceSystems.pdf

Before you begin: make sure you are not working withing a zip file. Unzip the files first.
- download test script: https://drive.google.com/file/d/19y6-R_mob8Nij-ItWW5Iaf7ulvqFVP_w/view?usp=sharing

Setting up your environment:
1. In the lower right panel on RStudio, navigate over to the InstallTest folder
2. Click on "More" > "Set as Working Directory"
Now your working directory has been correctly set (hopefully)
Note: ensure that the unzipped folder of data in also in this directory

# Ep.2 Supplemental Code

```
ggplot()+
  geom_histogram(data = DSM_HARV_df, aes=(HARV_dsmCrop))
```

```
#####################################
```

## Day 1 Questions:

Q: Could your working directory be a folder where I store my R projects?
A: Yes, your working directory can be wherever you store scripts and data (or the project you're using)

Q: ggplot2 and dplyr are both in the tidyverse right?
A: yes, they are both in the tidyverse

Q: Why scale_fill_viridis, rather than defaults?
A: nice pallet that is also color-blind friendly
A: scale_fill_virids provides easy-to-digest, colorblind friendly color palettes (more info here: https://cran.r-project.org/web/packages/viridis/vignettes/intro-to-viridis.html)

Q: what is a raster?
A: rasters = arrays of pixels
A: Raster = a matrix of pixels
A: Raster data: https://desktop.arcgis.com/en/arcmap/latest/manage-data/raster-and-images/what-is-raster-data.htm

Q: why do we chose HARV_dsmCrop as the fill aesthetic?
A: The HARV_dsmCrop is a column that contains the values of the elevation associated with each pixel. The values in this column are color-coded to create the colors seen in the map.

# Interval notation:

# (a,b] is the interval from a to b, not including a
# [a,b) is the interval from a to b, not including b
# (a,b) is the interval from a to b, not including either a or b
# [a,b] is the interval from a to b, inclusive of a and b

Q: Is terrain.colors() value going to be the same for every geom_raster() if there are three breaks?
A: Yes, you can expand the color scale using more breaks, but it will be the same set of colors every time

Q: What does guide = "none" mean?
A: This means "no legend"

Q: What's coord_quickmap()?
A: This creates a spatially relevant map; it tells ggplot() to do a rough plot of the coordinate reference system and makes the spatial scales accurate (avoids stretching of the map)

Q: What would happen if you passed scale_fill_viridis_c() before the 2nd geom_raster() argument?
A: It shouldn't change the output at all. The layer order matters because it plots one after another, but most of everything else can be added at any position

Q: How does a digital elevation model differ from a DSM or DTM?
A: DSM is the first return from a lidar source and includes all bumps and humps like trees and houses but not the actual ground. DTM shows the bare earth and represents lature returns from a lidar source and does not include trees and houses. These are both types of DEMs (digital elevation models)

Q: Why is terrain.colors(10) used? Is it because there are 10 breaks in the data we are plotting?
A: The number represents how many colors to give to the plot. Using 10 instead of 3 would change the specific colors shown, but it will use 10 colors to create a color ramp instead of 3. Here is a nice graphic at the bottom of the page: http://applied-r.com/color-ramp-functions/

Q: What are the issues associated with the projectRaster method?
A: projectRaster() is commonly used to reproject raster objects, but there are other data types that we might need to reproject that projectRaster doesn't work for

Q: What package is projectRaster() in?
A: It is in the raster package

Q: What is the difference between scale_fill_gradient() and scale_fill_gradientn()?
A: scale_fill_gradient() just takes a high and a low value color and creates its own color ramp, scale_fill_gradientn() lets you specify many different colors to create your color ramp

####################################

## Day 2 Questions:

Review:
Review.r file: https://drive.google.com/file/d/1PHUjBleQ1Ko4pa7yjABd21h4m-eLtiX5/view?usp=sharing

Q: What is a canopy height model?
A: It's the vertical difference between the canopy height (the height of the tree) and the ground level (the terrain model)
A2: The dataset we're using is a lidar dataset, which is where lasers and shot out of a plane. the lasers measure different 'returns' depening on whats on the ground.
the first return is what the lidar laser hits first which tends to be trees/buildings/other things between the machine and the ground. The last return is the bare ground.By manipulating the return data, you can make a digital terrain model (bare ground), a digital surface model, and canopy height model.

Q: Why don't we use aes() for plotting shapefiles?
A: In this example (ep 06), we don't use aes() because we're not calling to particular columns for our colors or fills; we want the entire object plotted with a cyan fill and a black bounding color

## Code for Episode 7:

```
lines_HARV <- st_read("data/NEON-DS-Site-Layout-Files/HARV/HARV_roads.shp")
point_HARV <- st_read("data/NEON-DS-Site-Layout-Files/HARV/HARVtower_UTM18N.shp")

footpath_HARV <- lines_HARV %>%
  filter(TYPE == "footpath")
```

```
ggplot()+
  geom_sf(data = footpath_HARV,
          aes(color = factor(OBJECTID)),
          size = 1.5)+
  ggtitle("NEON Harvard Forest Field Site",
          subtitle = "footpaths") +
  coord_sf()
```

Challenge; Make a plot with the stonewalls

```
stonewall_HARV <- lines_HARV %>%
  filter(TYPE == "stone wall")

ggplot()+
  geom_sf(data = stonewall_HARV,
          aes(color = factor(OBJECTID)),
          size = 1.5)+
  ggtitle("NEON Harvard Forest Field Site",
          subtitle = "Stone walls") +
  coord_sf()
```

## Code for Episode 8:

```
ggplot()+
  geom_sf(data = aoi_boundary_HARV,
          fill="grey",
          color = "gray") +
  geom_sf(data = lines_HARV,
          aes(color = TYPE),
          size = 1)
```

Bring up help screen with options of shapes you may choose from (default r shapes):

```
?pch
```

```
# Challenge
plot_locations <-
st_read("data/NEON-DS-Site-Layout-Files/HARV/PlotLocations_HARV.shp")
```

Q: How does scale_color_manual(values = road_color) know what sf to color?
A: The scale_color_manual() is looking for something specified in the geom_raster( aes(color = xx)).
So because Gavin wrote geom_sf(data = lines_HARV, aes(color = TYPE)), the scale_color_manual() will use that
Alternatively, Gavin wrote geom_sf(data = point_HARV, aes(fill = Sub_Type)), so the scale_fill_manual() is looking for things with fill = in the aes()
Ggplot is pretty smart!
scale_color_manual() only points to color in aes() and scale_fill_manual() only points to fill in aes()
So it looks for those attributes and assigns things accordingly

Q: What would you have to do differently if you had multiple shape files with an aes fill?

A:  Without loading additional packages that we won't go over today, you can only us aes fill with one of your shapefiles. If you try to use it on multiple shape files, it will only use it on the last shapefile you add to your plot. However, you can us a combination of aes color, aes fill, aes shape on different shapefiles. So part of the trick of making plots when you have multiple shapefiles is figuring out the aesthetics you want to map for each one.

Challenge Question:
1. Using the NEON-DS-Site-Layout-Files/HARV/PlotLocations_HARV.shp shapefile, create a map of study plot locations, with each point colored by the soil type (soilTypeOr). How many different soil types are there at this particular field site? Overlay this layer on top of the lines_HARV layer (the roads). Create a custom legend that applies line symbols to lines and point symbols to the points.
2. Modify the plot above. Tell R to plot each point, using a different symbol of shape value.

Great R color cheatsheet by NCEAS
https://www.nceas.ucsb.edu/sites/default/files/2020-04/colorPaletteCheatsheet.pdf

## Code for Episode 9: CRS

```
state_boundary_US <- st_read("data/NEON-DS-Site-Layout-Files/US-Boundary-Layers/US-State-Boundaries-Census-2014.shp")

#plot states boundaries
ggplot() +
  geom_sf(data = state_boundary_US) +
  ggtitle("Map of Contiguous US State Boundaries") +
  coord_sf()

NE.States.Boundary.US <- st_read("data/NEON-DS-Site-Layout-Files/US-Boundary-Layers/Boundary-US-State-NEast.shp")
```

Episode 10

```
plot_locations_HARV <-
read.csv("data/NEON-DS-Site-Layout-Files/HARV/HARV_PlotLocations.csv")
```

Drivers: links to full driver documentation are found at https://gdal.org/drivers/vector/index.html