Welcome to Etherpad!

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

Get involved with Etherpad at http://etherpad.org

*Attendance April 19, 2019: https://forms.gle/8nBTqK9GBKuCYNcy8

*Guest wifi ID and Password:
*   Choose CHSguest, no password needed

- Download spreadsheet 1:   https://ndownloader.figshare.com/files/2252083
- Quality control exercise spreadsheet: https://github.com/datacarpentry/spreadsheet-ecology-lesson/blob/gh-pages/data/survey_sorting_exercise.xlsx?raw=true

*
*
*OpenRefine Data: https://ndownloader.figshare.com/files/7823341

Clustering In Depth: https://github.com/OpenRefine/OpenRefine/wiki/Clustering-In-Depth

1+1
sqrt(9)
??sqrt #for help use ??
weight_kg <- 55 #this creates a new object
weight_kg*2.2
weight_kg <- 57.5
weight_lb<-weight_kg*2.2 #can now use weight in lbs as new object
weight_kg <- 55
mass<- 50
age<- 120
mass<-mass*2
age<-age-20
mass_index<-mass/age

Download the data for the R workshop

download.file("https://ndownloader.figshare.com/files/2292169", "data/portal_data_joined.csv")

*---
-----

*R—Day 02
download.file("https://ndownloader.figshare.com/files/2292169", "data/portal_data_joined.csv")

install.packages("tidyverse")

library("tidyverse")

surveys <- read_csv("data/portal_data_joined.csv")

surveys %>% group_by(sex) %>% summarize(mean_weight = mean(weight, na.rm = TRUE))

surveys %>% group_by(sex, species_id) %>% summarize(mean_weight = mean(weight, na.rm = TRUE))

surveys %>% group_by(sex, species_id) %>% summarize(mean_weight = mean(weight, na.rm = TRUE)) %>% print(n = 92)

surveys %>% filter(!is.na(weight)) %>% group_by(sex, species_id) %>% summarize(mean_weight = mean(weight)) %>% print(n = 92)

surveys %>% filter(!is.na(weight)) %>% group_by(sex, species_id) %>% summarize(mean_weight = mean(weight), min_weight = min(weight))

surveys %>% group_by(sex) %>% tally()

surveys_complete <- surveys %>% filter(species_id!= "", !is.na(weight), !is.na(hindfoot_length), sex! ="")

species_counts <- surveys_complete %>% group_by(species_id) %>% tally() %>% filter(n >= 50)

species_counts

surveys_complete <- surveys_complete %>% filter(species_id %in% species_counts$species_id)

dim(surveys_complete)
# output should show something like "30463   13"

My code so far...

library("tidyverse")

```
surveys <- read_csv("data/portal_data_joined.csv")

surveys %>%
  group_by(sex) %>%
  summarize(mean_weight = mean(weight,
                    na.rm = TRUE))

surveys %>%
  group_by(sex, species_id) %>%
  summarize(mean_weight = mean(weight,
                    na.rm = TRUE))

surveys %>%
  group_by(sex, species_id) %>%
  summarize(mean_weight = mean(weight,
                    na.rm = TRUE)) %>%
  print(n = 92)

surveys %>%
  filter(!is.na(weight)) %>%
  group_by(sex, species_id) %>%
  summarise(mean_weight = mean(weight)) %>%
  print(n = 92)

surveys %>%
  filter(!is.na(weight)) %>%
  group_by(sex, species_id) %>%
  summarise(mean_weight = mean(weight),
        min_weight = min(weight))

surveys %>%
  group_by(sex) %>%
  tally()

surveys_complete <- surveys %>%
  filter(species_id != "",
       !is.na(weight),
       !is.na(hindfoot_length),
       sex != "")

species_counts <- surveys_complete %>%
  group_by(species_id) %>%
  tally() %>%
  filter(n >= 50)

species_counts

surveys_complete <- surveys_complete %>%
  filter(species_id %in% species_counts$species_id)

dim(surveys_complete)


ggplot(data = surveys_complete)
```

```r
ggplot(data = surveys_complete, aes(x = weight, y = hindfoot_length))

ggplot(data = surveys_complete, aes(x = weight, y = hindfoot_length)) + geom_point()

ggplot(data = surveys_complete, aes(x = weight, y = hindfoot_length)) + geom_point(alpha = 0.1)

surveys_plot <- ggplot(data = surveys_complete, aes(x = weight, y = hindfoot_length))

surveys_plot

surveys_plot + geom_point(alpha = 0.1, color = "blue")

surveys_plot + geom_point(alpha = 0.1, aes(color = species_id))

surveys_plot <- ggplot(data = surveys_complete, aes(x = species_id, y = weight))

surveys_plot + geom_boxplot()

surveys_plot + geom_boxplot(alpha = 0) + geom_jitter(alpha = 0.3, color = "tomato")

surveys_plot + geom_jitter(alpha = 0.3, color = "tomato") + geom_boxplot(alpha = 0)

yearly_counts <- surveys_complete %>% group_by(year, species_id) %>% tally()

timelapse <- ggplot(data = yearly_counts, aes(x = year, y = n))

timelapse + geom_line()

timelapse <- ggplot(data = yearly_counts, aes(x = year, y = n, group = species_id))

timelapse + geom_line()

timelapse <- ggplot(data = yearly_counts, aes(x = year, y = n,  color = species_id))

timelapse + geom_line()

timelapse <- ggplot(data = yearly_counts, aes(x = year, y = n))

timelapse + geom_line() + facet_wrap(~ species_id)

timelapse + geom_line() + facet_wrap(~ species_id) + theme_bw()

timelapse + geom_line() + facet_wrap(~ species_id) + theme_bw() + theme(panel.grid = element_blank())

install.packages("gridExtra")
library("gridExtra")




count_plot <- ggplot(data = yearly_counts, aes(x = year, y = n, color = species_id)) + geom_line() + xlab("Year") + ylab("Abundance")

count_plot
```

```
combo_plot <- grid.arrange(weight_boxplot, count_plot, ncol = 2, widths = c(4,6))

ggsave("combo_plot.png", combo_plot, width = 10, dpi = 300)
```

*SQL

Link for the dataset: https://doi.org/10.6084/m9.figshare.1314459

*Contact Information:
   Phillip Doehle—doehle@okstate.edu
   Kay Bjornen—kay.bjornen@okstate.edu
   Ki Cole—ki.cole@okstate.edu

helper - Michael Anderson - michael.b.anderson@okstate.edu

helper - Peter Hoyt (peter.r.hoyt@okstate.edu)

*Additional Resources

- Data Carpentry (Workshop lessons)—http://www.datacarpentry.org/
- Software Carpentry (Similar lessons)—https://software-carpentry.org/
- R for Data Science (Good next step)—http://r4ds.had.co.nz/
- Tidyverse Website (Documentation and Cheet Sheets for Tidyverse Packages)— https://www.tidyverse.org/
- Tidy Data by Hadley Wickham (Paper on the Tidy Data philosophy which informs development of Tidyverse software)—http://vita.had.co.nz/papers/tidy-data.html