

02-data-project

Kurt Ingeman

9/20/2021

This is going to be a chunk of code

read in data and merge with spp list

```
df <- read.csv(here("Data", "Marianas_fishabund_numperm2_sector.csv")) %>%
  rename(SECTOR = ANALYSIS_SEC)

sum(df$N) # 1217 surveys

## [1] 1217

# remove standard error columns for reshape
df_wide <- df %>%
  select(!contains("_SE"))

# reshape data
df_long <- df_wide %>%
  pivot_longer(cols = 8:500, names_to = "SPECIES", values_to = "DENSITY") %>%
  filter(DENSITY > 0)

# How many surveys?
length(unique(df_long$SPECIES)) #471

## [1] 471

# read in species list
spp <- read.csv(here("Data", "FISH_SPECIES_04-21-2020.csv")) %>%
  select(SPECIES, TAXONNAME, FAMILY, TROPHIC, TROPHIC_SIMPLE, TROPHIC_MONREP)

# merge with species list to incorporate trophic groups
df_long <- left_join(df_long, spp, by = "SPECIES") %>%
  select(-X)
```

data summaries

```
richness_by_yr <- df_long %>%
  group_by(ANALYSIS_YEAR) %>%
  summarise(RICHNESS = length(unique(SPECIES)))

richness_by_yr %>% gt() %>%
  tab_header(
```

ANALYSIS_YEAR	RICHNESS	ANNUAL_SURVEYS	spp_per_survey
2009	337	174	1.936782
2011	402	354	1.135593
2014	366	358	1.022346
2017	381	331	1.151057

```
title = "Species Richness",
subtitle = "By Year")
```

Species Richness

By Year

ANALYSIS_YEAR	RICHNESS
2009	337
2011	402
2014	366
2017	381

```
surveys_by_year <- df_long %>%
  group_by(ANALYSIS_YEAR, SECTOR) %>%
  summarise(N = N) %>%
  slice(1) %>%
  group_by(ANALYSIS_YEAR) %>%
  summarise(ANNUAL_SURVEYS = sum(N))
```

`summarise()` has grouped output by 'ANALYSIS_YEAR', 'SECTOR'. You can override using the `groups`

```
richness_surveys_year <- left_join(richness_by_yr, surveys_by_year, by = "ANALYSIS_YEAR") %>%
  group_by(ANALYSIS_YEAR) %>%
  mutate(spp_per_survey = RICHNESS / ANNUAL_SURVEYS)
```

```
richness_surveys_year %>%
  kbl() %>%
  kable_classic_2(full_width = F)
```

```
df_long <- left_join(df_long, richness_surveys_year, by = "ANALYSIS_YEAR")
```

```
richness_sector <- df_long %>%
  group_by(ANALYSIS_YEAR, SECTOR) %>%
  mutate(SURVEYS = N,
         RICHNESS = length(unique(SPECIES)),
         RICH_DENSE = RICHNESS / SURVEYS) %>%
  slice(1) %>%
  select(SECTOR, ANALYSIS_YEAR, SURVEYS, RICHNESS, RICH_DENSE) %>%
  arrange(ANALYSIS_YEAR, desc(RICH_DENSE))
```

```
richness_sector %>%
  gt(groupname_col = "SECTOR") %>%
  tab_header(
    title = "Fish Species Richness",
```

```
subtitle = "By Year in each sector")
```

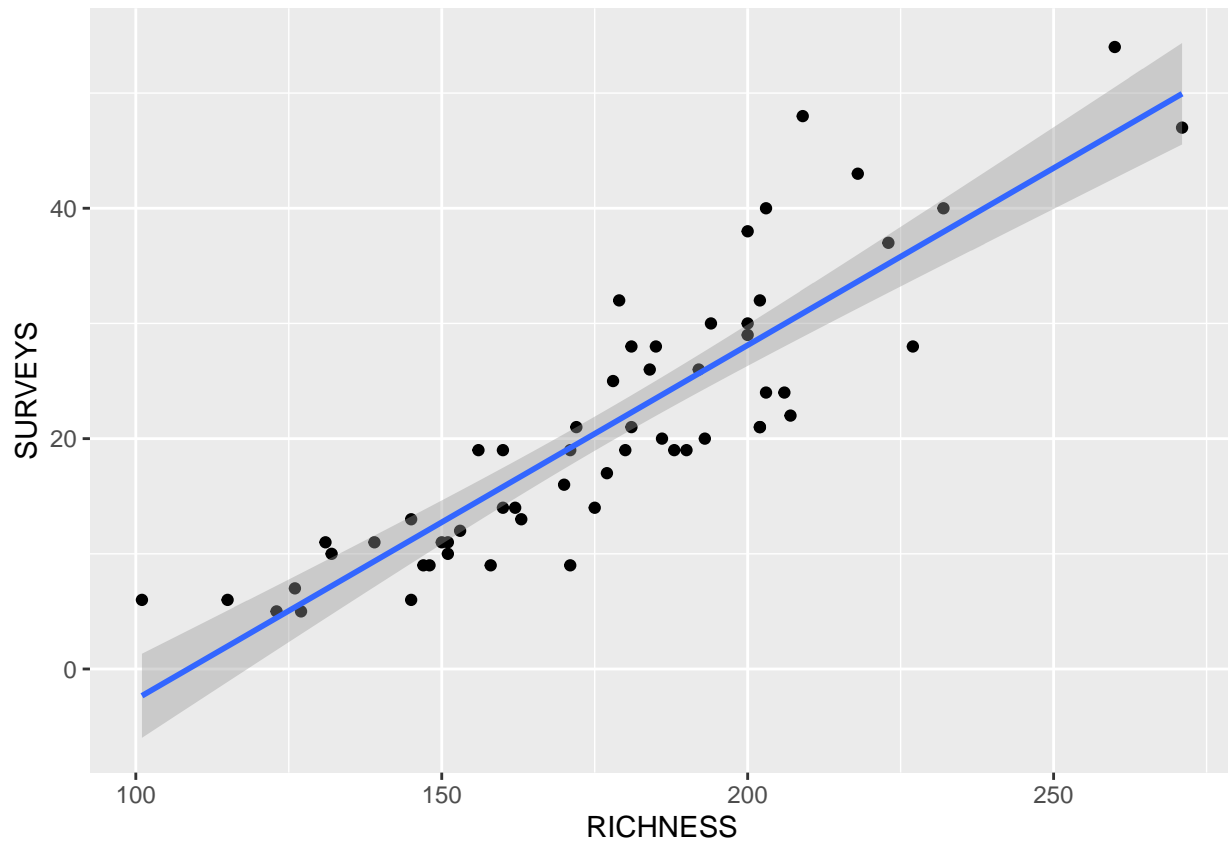
Fish Species Richness				
By Year in each sector				
	ANALYSIS_YEAR	SURVEYS	RICHNESS	RICH_DENSE
Guguan				
	2009	5	127	25.400000
	2011	10	151	15.100000
	2014	11	139	12.636364
	2017	9	147	16.333333
Alamagan				
	2009	6	145	24.166667
	2011	5	123	24.600000
	2014	11	151	13.727273
	2017	9	148	16.444444
Aguijan				
	2009	6	115	19.166667
	2011	13	145	11.153846
	2014	10	132	13.200000
	2017	17	177	10.411765
Sarigan				
	2009	7	126	18.000000
	2011	9	171	19.000000
	2014	11	150	13.636364
	2017	9	158	17.555556
Farallon de Pajaros				
	2009	6	101	16.833333
	2011	12	153	12.750000
	2014	11	131	11.909091
	2017	16	170	10.625000
Asuncion				
	2009	13	163	12.538462
	2011	20	186	9.300000
	2014	21	172	8.190476
	2017	19	171	9.000000
Rota				
	2009	14	175	12.500000
	2011	24	203	8.458333
	2014	28	185	6.607143
	2017	28	227	8.107143
Agrihan				
	2009	14	162	11.571429
	2011	20	193	9.650000
	2017	19	188	9.894737
Tinian				
	2009	14	160	11.428571

	2011	19	190	10.000000
	2014	19	160	8.421053
	2017	24	206	8.583333
Pagan				
	2009	21	202	9.619048
	2011	29	200	6.896552
	2014	43	218	5.069767
	2017	40	232	5.800000
Saipan				
	2009	22	207	9.409091
	2011	30	194	6.466667
	2014	48	209	4.354167
	2017	37	223	6.027027
Maug				
	2009	21	181	8.619048
	2011	30	200	6.666667
	2014	40	203	5.075000
	2017	38	200	5.263158
Guam				
	2009	25	178	7.120000
GUA_EAST_OPEN				
	2011	32	202	6.312500
	2014	28	181	6.464286
	2017	19	156	8.210526
GUA_MP				
	2011	47	271	5.765957
	2017	21	202	9.619048
GUA_WEST_OPEN				
	2011	54	260	4.814815
	2014	26	184	7.076923
	2017	26	192	7.384615
GUA_MP_MINUS_ACHANG				
	2014	19	180	9.473684
GUA_ACHANG				
	2014	32	179	5.593750

data visulization

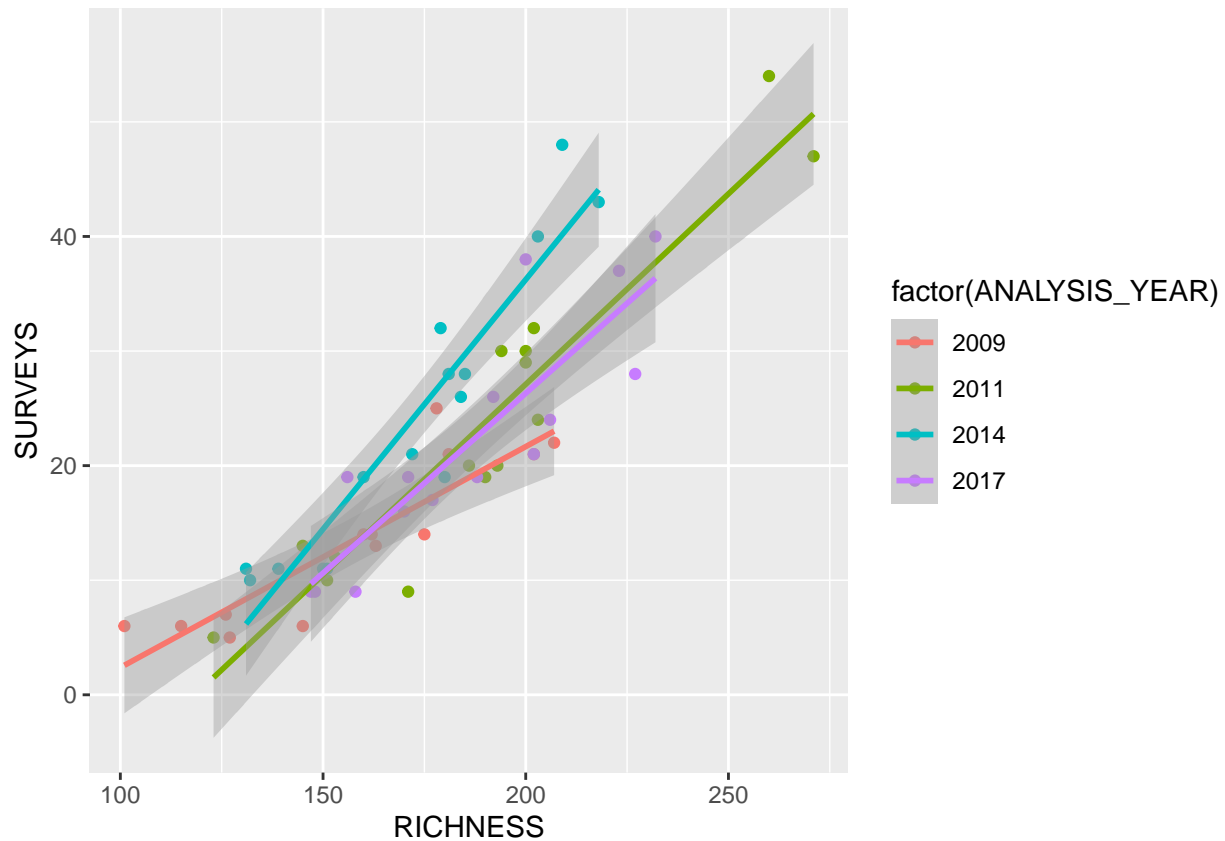
```
ggplot(data = richness_sector, aes(x = RICHNESS, y = SURVEYS)) +
  geom_point() +
  geom_smooth(method = "lm")
```

```
## `geom_smooth()` using formula 'y ~ x'
```



```
ggplot(data = richness_sector, aes(x = RICHNESS, y = SURVEYS, color = factor(ANALYSIS_YEAR))) +  
  geom_point() +  
  geom_smooth(method = "lm")
```

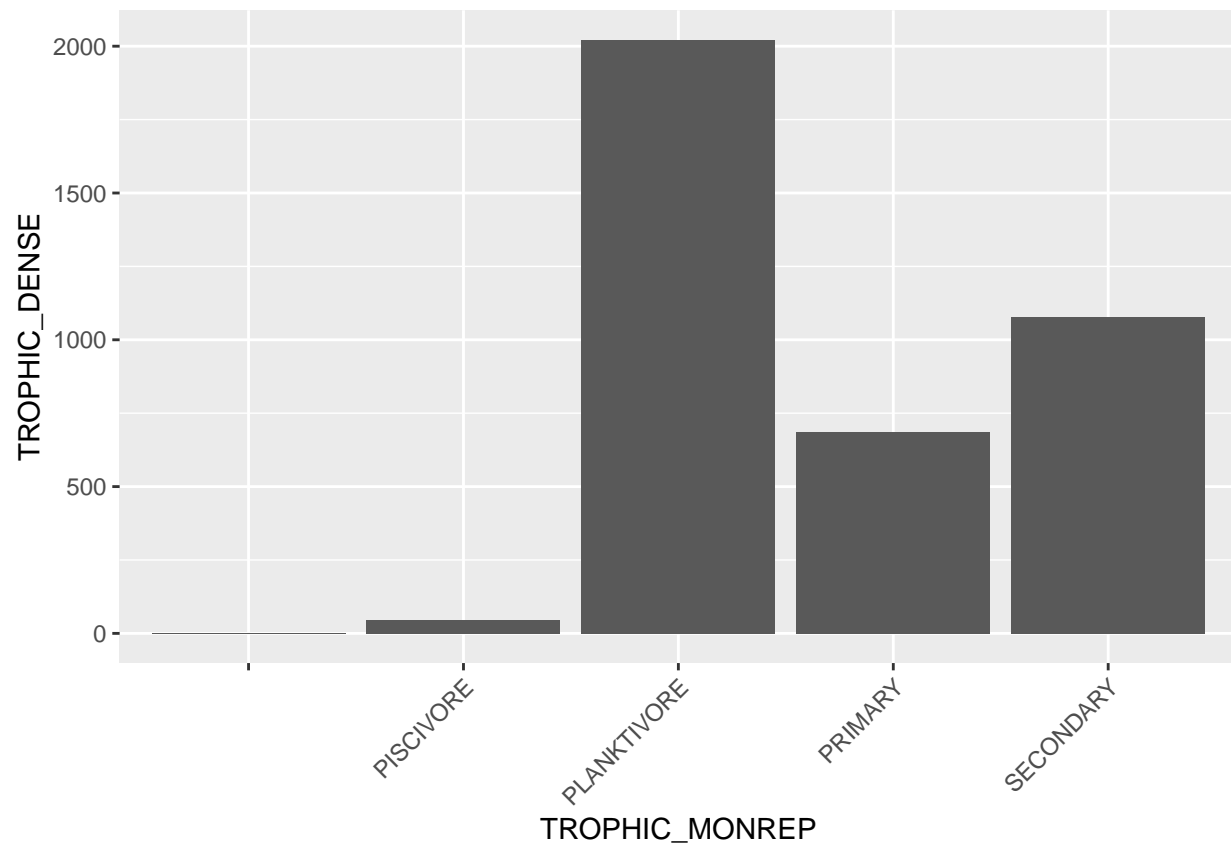
```
## `geom_smooth()` using formula 'y ~ x'
```



data calculations

```
df_troph <- df_long %>%
  group_by(ANALYSIS_YEAR, SECTOR, TROPHIC_MONREP) %>%
  mutate(TROPHIC_DENSE = sum(DENSITY)) %>%
  mutate(YEAR = factor(ANALYSIS_YEAR),
         SECTOR = factor(SECTOR)) %>%
  ungroup()

ggplot(df_troph, aes(x = TROPHIC_MONREP, y = TROPHIC_DENSE)) +
  geom_bar(stat = "identity") +
  theme(axis.text.x = element_text(angle = 45, hjust=1))
```

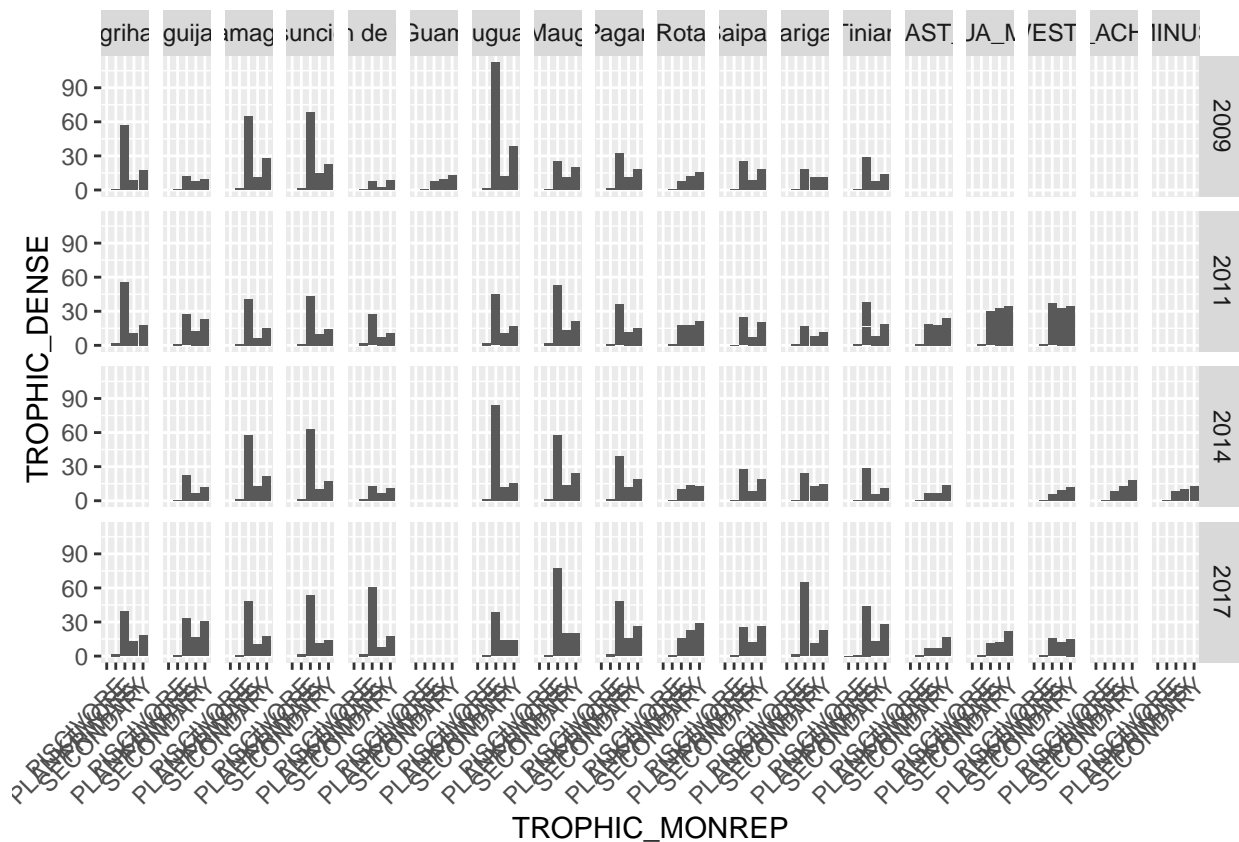


Facet by year and by sector

```
ggplot(df_troph, aes(x = TROPIC_MONREP, y = TROPIC_DENSE)) +  
  geom_bar(stat = "identity") +  
  facet_grid( vars(YEAR), vars(SECTOR)) +  
  theme(axis.text.x = element_text(angle = 45, hjust=1))
```

	mpg	cyl	disp	hp	drat	wt
Mazda RX4	21.0	6	160	110	3.90	2.620
Mazda RX4 Wag	21.0	6	160	110	3.90	2.875
Datsun 710	22.8	4	108	93	3.85	2.320
Hornet 4 Drive	21.4	6	258	110	3.08	3.215
Hornet Sportabout	18.7	8	360	175	3.15	3.440

	mpg	cyl	disp	hp	drat	wt
Mazda RX4	21.0	6	160	110	3.90	2.620
Mazda RX4 Wag	21.0	6	160	110	3.90	2.875
Datsun 710	22.8	4	108	93	3.85	2.320
Hornet 4 Drive	21.4	6	258	110	3.08	3.215
Hornet Sportabout	18.7	8	360	175	3.15	3.440



```
library(kableExtra)
dt <- mtcars[1:5, 1:6]
```

```
dt %>%
  kbl() %>%
  kable_styling()
```

```
dt %>%
  kbl() %>%
  kable_classic_2(full_width = F)
```


	mpg	cyl	disp	hp	drat	wt
Mazda RX4	21.0	6	160	110	3.90	2.620
Mazda RX4 Wag	21.0	6	160	110	3.90	2.875
Datsun 710	22.8	4	108	93	3.85	2.320
Hornet 4 Drive	21.4	6	258	110	3.08	3.215
Hornet Sportabout	18.7	8	360	175	3.15	3.440

```
dt %>%
  kbl() %>%
  kable_paper("hover", full_width = F)
```