

TC data analysis – SI and SP basins

Melissa Renard

2025-02-19

```
# CHOOSE BASIN HERE -- 'SI' OR 'SP'.
basin_choice <- 'SI'

if (basin_choice == 'SI'){
  x_limits <- c(30, 140)
  x_breaks <- seq(30,140,10)
  y_limits <- c(-40, 0)
  y_breaks <- seq(-40,0,10)
} else if (basin_choice == 'SP'){
  x_limits <- c(130, 240)
  x_breaks <- seq(130,240,10)
  y_limits <- c(-40, 0)
  y_breaks <- seq(-40,0,10)
} else {
  stop("basin_choice must be 'SI' or 'SP'.")
}

# Load tracks data
tracks <- read.csv("../Data/tracks.csv", na="") %>% as_tibble %>%
  filter(basin == basin_choice)

tracks$category <- tracks$category %>% as.factor
tracks <- tracks %>% rename('t_day' = 'days_since_1980',
  't_year' = 'years_since_1980',
  'Nt' = 'TC_num') %>%
  mutate(t_month = lapply(t_year, function(x) x*12) %>% unlist,
    east170 = lon > 170,
    east75 = lon > 75,
    range_sw = max_sw - min_sw,
    Nt = as.integer(rownames(tracks)))

# TCs in the southern hemisphere occur between November and April.
# Need to write a new dataframe for that.
tracks_1 <- tracks %>% mutate(enso_year = year + (month > 7)) %>%
  filter(enso_year >= 1981 & enso_year <= 2023)

# Load ENSO data
enso_phases <- c("Nina", "Neutral", "Nino")
enso_labels <- c("La Niña", "Neutral", "El Niño")

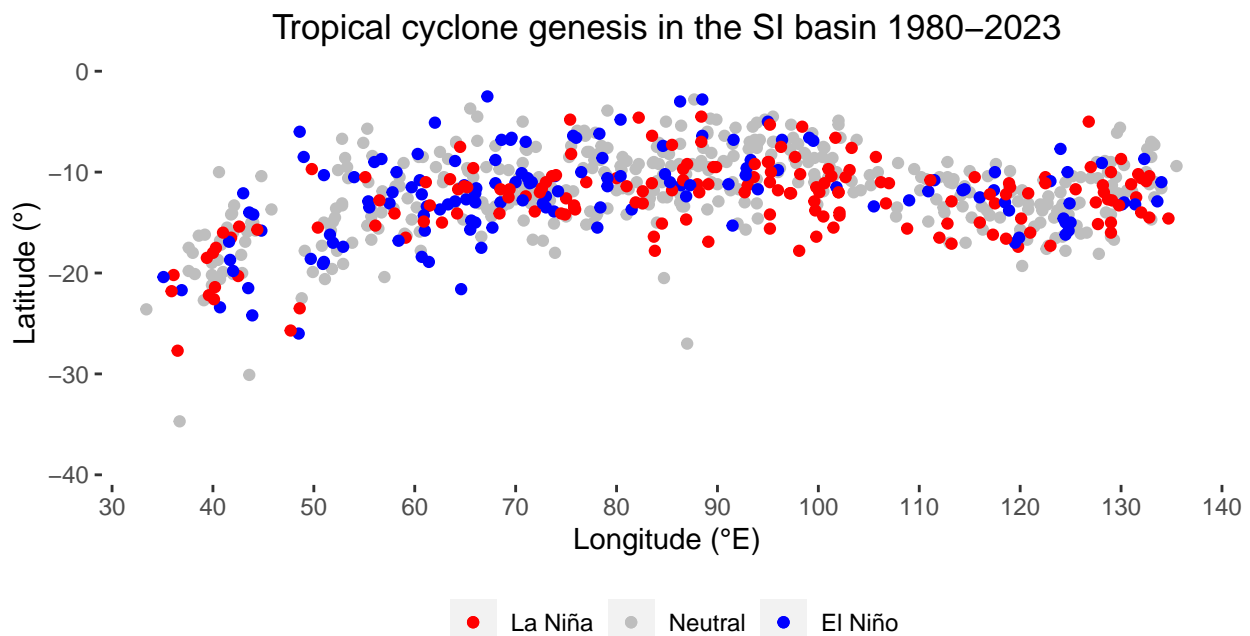
enso_df <- read.csv("../Data/ENSO.csv", skip=3) %>% as_tibble %>%
  mutate(enso = factor(enso, levels = enso_phases))
enso_df_TC <- enso_df %>% filter(year >= 1981 & year <= 2023)
```

```

tracks_1 <- tracks_1 %>% left_join(enso_df, by = join_by(enso_year==year))

ggplot(tracks_1, aes(x = lon, y = lat, color = enso)) +
  geom_point(data = subset(tracks_1, enso== 'Neutral'), aes(color = enso), size = 1.5)+
  geom_point(data = subset(tracks_1, enso!= 'Neutral'), aes(color = enso), size = 1.5) +
  scale_x_continuous(limits = x_limits, breaks = x_breaks, expand = c(0,1)) +
  scale_y_continuous(limits = y_limits, breaks = y_breaks, expand=c(0,1)) +
  labs(
    title = "Tropical cyclone genesis in the SI basin 1980-2023",
    x = "Longitude (\u00B0E)",
    y = "Latitude (\u00B0)",
    color = "ENSO Phase"
  ) +
  coord_equal() + # Ensure equal scaling of x and y axes
  theme(
    plot.title = element_text(hjust = 0.5),
    panel.border = element_blank(),
    panel.grid = element_blank(),
    panel.background = element_blank(),
    legend.position = 'bottom',
  ) +
  scale_color_manual(
    name = "",
    breaks = enso_phases,
    values = c("red", "grey", "blue"),
    labels = enso_labels
  )
)

```



```

p1 <- ggplot(tracks, aes(x = year)) +
  geom_bar() +
  theme_classic() +
  labs(x = "Year", y = "Count", title = "Number of TCs in SP basin\nby year") +
  theme(plot.title = element_text(hjust = 0.5))

```

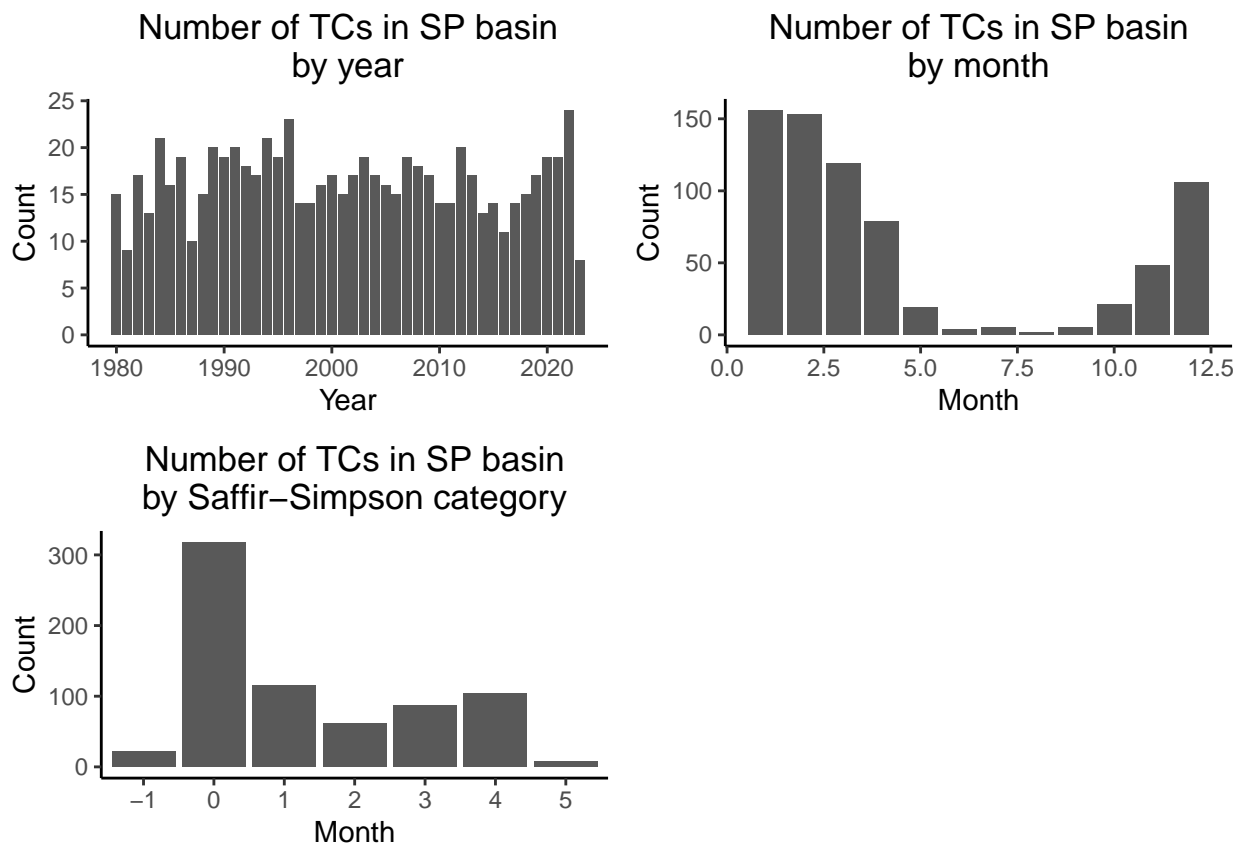
```

p2 <- ggplot(tracks_1, aes(x = month)) +
  geom_bar() +
  theme_classic() +
  labs(x = "Month", y = "Count", title = "Number of TCs in SP basin\nby month") +
  theme(plot.title = element_text(hjust = 0.5))

p3 <- ggplot(tracks_1, aes(x = category)) +
  geom_bar() +
  theme_classic() +
  labs(x = "Month", y = "Count", title = "Number of TCs in SP basin\nby Saffir-Simpson category") +
  theme(plot.title = element_text(hjust = 0.5))

grid.arrange(p1,p2,p3, ncol=2, nrow=2)

```



Frequency

Frequency based on ENSO phases

```

my_lm <- glm(table(tracks_1$enso_year) ~ 1, family = poisson)
summary(my_lm)

##
## Call:
## glm(formula = table(tracks_1$enso_year) ~ 1, family = poisson)
##
## Deviance Residuals:

```

```
##      Min      1Q      Median      3Q      Max
## -2.36599 -0.67374  0.07947  0.55699  1.68208
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  2.81388    0.03735   75.35  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##      Null deviance: 36.773  on 42  degrees of freedom
## Residual deviance: 36.773  on 42  degrees of freedom
## AIC: 238.07
##
## Number of Fisher Scoring iterations: 4
mu <- my_lm$coefficients[[1]]
exp(mu)

## [1] 16.67442
my_lm <- glm(table(tracks_1$enso_year) ~ enso_df_TC$enso, family = poisson)

summary(my_lm)

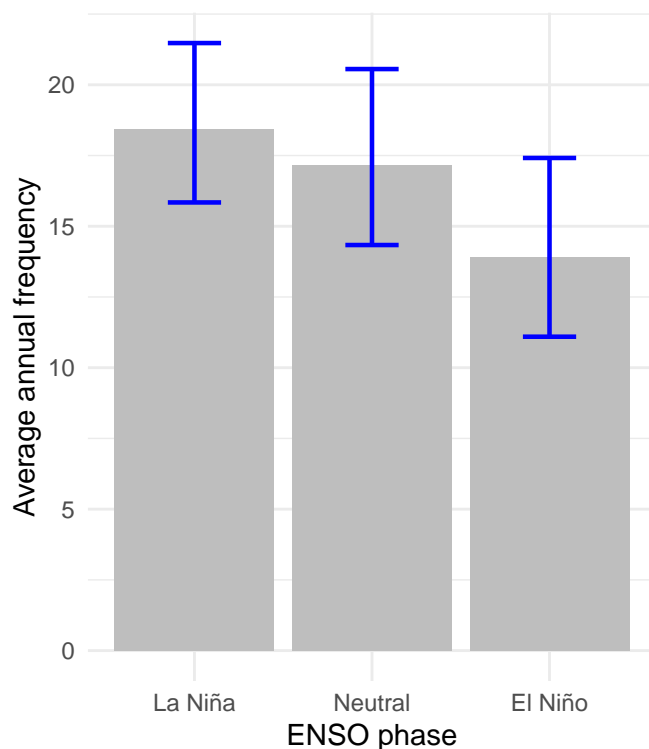
##
## Call:
## glm(formula = table(tracks_1$enso_year) ~ enso_df_TC$enso, family = poisson)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.72071  -0.53456   0.02679   0.56598   1.55479
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)    2.91476    0.07762  37.554  <2e-16 ***
## enso_df_TC$ensoNeutral -0.07179    0.09193  -0.781   0.4348
## enso_df_TC$ensoNino   -0.28287    0.11497  -2.460   0.0139 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##      Null deviance: 36.773  on 42  degrees of freedom
## Residual deviance: 29.898  on 40  degrees of freedom
## AIC: 235.19
##
## Number of Fisher Scoring iterations: 4
mu <- c(my_lm$coefficients[[1]],
        my_lm$coefficients[[1]] + my_lm$coefficients[[2]],
        my_lm$coefficients[[1]] + my_lm$coefficients[[3]])
exp(mu)
```

```
## [1] 18.44444 17.16667 13.90000
```

Estimated mean frequency of TCs in SI basin in La Niña, neutral and El Niño phases, respectively: 18.44, 17.17, 13.9

```
tracks_1 %>%
  count(enso) %>%
  mutate(fqcy = n/c(9,24,10),
         sde = summary(my_lm)$coef[,2] %>% as.vector) %>%
  ggplot(aes(x = enso, y = fqcy)) +
  geom_bar(stat = 'identity', fill = 'grey') +
  geom_errorbar(aes(x = enso, ymin = fqcy*exp(1.96*sde),
                    ymax = fqcy*exp(-1.96*sde)), width = 0.3, color="blue",
               size=0.8)+
  theme_minimal() +
  labs(x = 'ENSO phase', y = 'Average annual frequency') +
  scale_x_discrete(
    labels = enso_labels
  ) +
  theme(legend.position = 'None')
```

```
## Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use `linewidth` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```



Locations frequency based on ENSO phases

```
summary(tracks_1$lon)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  33.40   65.60   85.50   86.55  110.10  135.50
```

```
summary(tracks_1$lat)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -34.70  -14.30  -11.80  -12.15   -9.60   -2.50
```

```
my_lm <- lm(lon ~ enso, data=tracks_1)
```

```
summary(my_lm)
```

```
##
## Call:
## lm(formula = lon ~ enso, data = tracks_1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -54.90 -18.93  -1.43   22.40   54.81
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   90.796      2.086  43.534 < 2e-16 ***
## ensoNeutral   -3.466      2.470  -1.403  0.160981
## ensoNino     -11.611      3.089  -3.758  0.000185 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 26.87 on 714 degrees of freedom
## Multiple R-squared:  0.02048,    Adjusted R-squared:  0.01774
## F-statistic: 7.466 on 2 and 714 DF,  p-value: 0.000618

mu <- c(my_lm$coefficients[[1]],
        my_lm$coefficients[[1]] + my_lm$coefficients[[2]],
        my_lm$coefficients[[1]] + my_lm$coefficients[[3]])

mu
```

```
## [1] 90.79639 87.32990 79.18561
```

Estimated mean longitude of TCs in La Niña, neutral and El Niño phases, respectively: 90.796, 87.33, 79.186.

This is significant. La Niña means more TCs close to Australia.

```
if (basin_choice == 'SI'){
  my_lm <- glm(east75 ~ enso, data=tracks_1, family = binomial)
} else {
  my_lm <- glm(east170 ~ enso, data=tracks_1, family = binomial)
}

summary(my_lm)
```

```
##
## Call:
## glm(formula = east75 ~ enso, family = binomial, data = tracks_1)
```

```
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.5753  -1.4452   0.8262   0.9314   1.2580
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)   0.8995     0.1712   5.254 1.49e-07 ***
## ensoNeutral  -0.2890     0.1999  -1.446   0.148
## ensoNino     -1.0871     0.2415  -4.501 6.77e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 948.81  on 716  degrees of freedom
## Residual deviance: 925.62  on 714  degrees of freedom
## AIC: 931.62
##
## Number of Fisher Scoring iterations: 4
mu <- c(my_lm$coefficients[[1]],
        my_lm$coefficients[[1]] + my_lm$coefficients[[2]],
        my_lm$coefficients[[1]] + my_lm$coefficients[[3]])
exp(mu)/(1+exp(mu))
```

```
## [1] 0.7108434 0.6480583 0.4532374
```

Estimated probability of a TC produced east of 75°E in La Niña, neutral and El Niño phases, respectively: 0.7108434, 0.6480583, 0.4532374

Intensity

Intensity based on ENSO phases

```
my_lm <- glm(max_sw ~ enso, data=tracks_1, family = Gamma(link=log))
summary(my_lm)

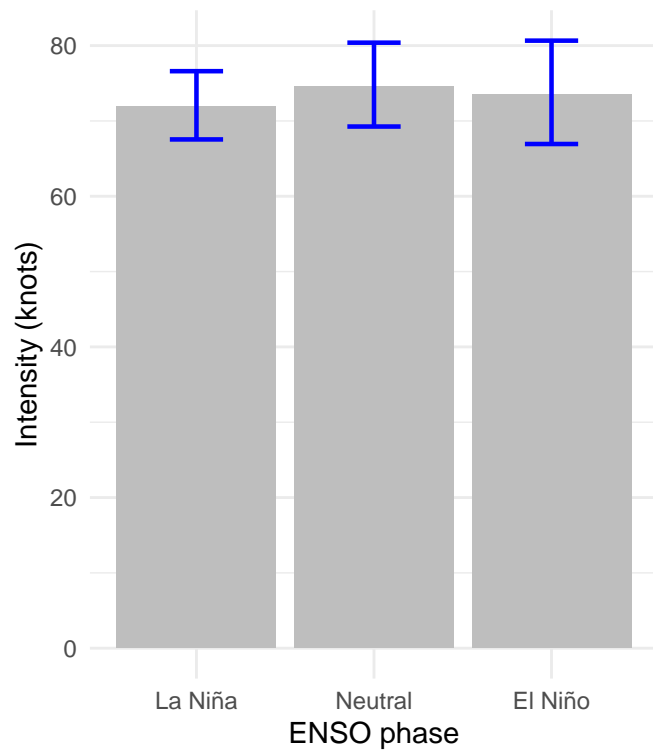
##
## Call:
## glm(formula = max_sw ~ enso, family = Gamma(link = log), data = tracks_1)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.99333  -0.37528  -0.08881   0.30997   0.83870
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   4.27573    0.03211 133.145 <2e-16 ***
## ensoNeutral   0.03659    0.03804   0.962   0.336
## ensoNino      0.02125    0.04757   0.447   0.655
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## (Dispersion parameter for Gamma family taken to be 0.1711899)
##
## Null deviance: 126.08 on 716 degrees of freedom
## Residual deviance: 125.92 on 714 degrees of freedom
## AIC: 6858.1
##
## Number of Fisher Scoring iterations: 4
mu <- c(my_lm$coefficients[[1]],
        my_lm$coefficients[[1]] + my_lm$coefficients[[2]],
        my_lm$coefficients[[1]] + my_lm$coefficients[[3]])

exp(mu)

## [1] 71.93264 74.61364 73.47776
sde <- summary(my_lm)$coef[,2] %>% as.vector

tracks_1 %>%
  count(enso) %>%
  mutate(sev = exp(mu),
         sde = summary(my_lm)$coef[,2] %>% as.vector) %>%
  ggplot(aes(x = enso, y = sev)) +
  geom_bar(stat = 'identity', fill = 'grey') +
  geom_errorbar(aes(x = enso, ymin = exp(mu - 1.96*sde),
                    ymax = exp(mu + 1.96*sde)), width = 0.3, color="blue",
               size=0.8)+
  theme_minimal() +
  labs(x = 'ENSO phase', y = 'Intensity (knots)') +
  scale_x_discrete(
    labels = enso_labels
  ) +
  theme(legend.position = 'None')
```

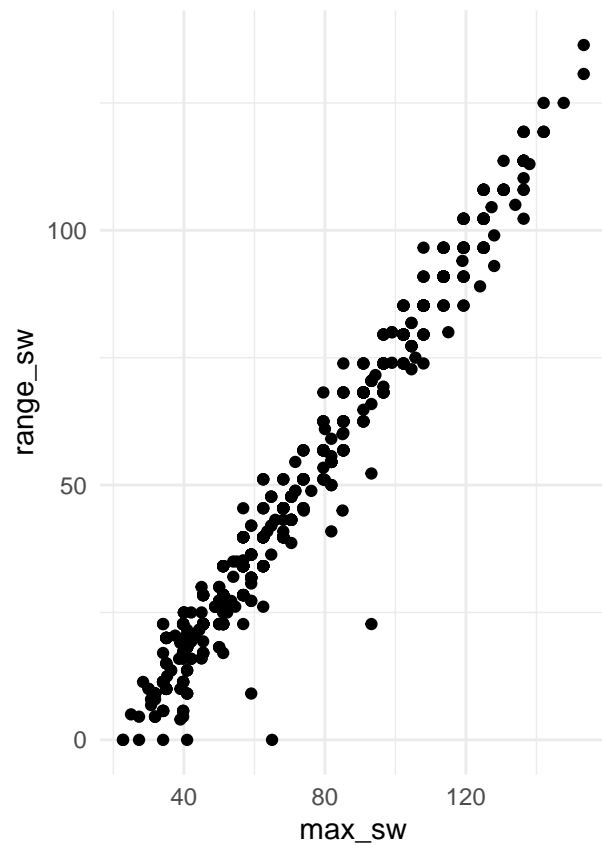
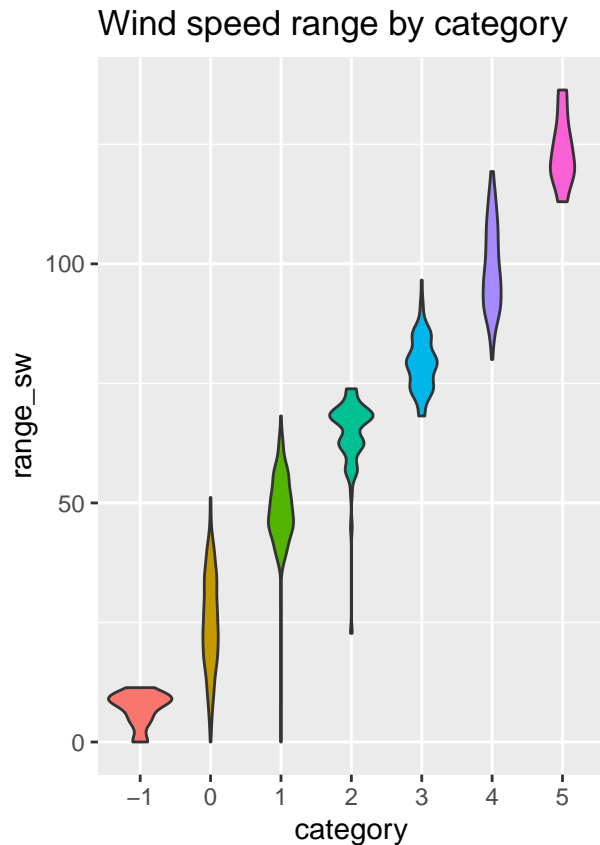



The estimated mean maximum wind speed of a TC in La Niña, Neutral and El Niño phases are, respectively: 0.23, 0.23, 0.23.

```
p1 <- ggplot(data = tracks_1, aes(x = category, y = range_sw)) +
  geom_violin(aes(fill = category)) +
  labs(title = "Wind speed range by category") +
  theme(legend.position = "none")

p2 <- ggplot(data = tracks_1, aes(x = max_sw, y = range_sw)) +
  geom_point() +
  theme_minimal()

grid.arrange(p1, p2, ncol = 2)
```



Track lifetime

```
my_lm <- glm(lifetime ~ enso, data = tracks_1, family=Gamma(link=log))
summary(my_lm)
```

```
##
## Call:
## glm(formula = lifetime ~ enso, family = Gamma(link = log), data = tracks_1)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -2.22622  -0.41900  -0.07079   0.25643   1.46760
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -3.814044   0.038884 -98.088  <2e-16 ***
## ensoNeutral  -0.026680   0.046056  -0.579    0.563
## ensoNino     -0.005188   0.057599  -0.090    0.928
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for Gamma family taken to be 0.2509849)
##
##      Null deviance: 218.96  on 716  degrees of freedom
## Residual deviance: 218.85  on 714  degrees of freedom
## AIC: -4483.3
```

```
##
## Number of Fisher Scoring iterations: 5
mu <- c(my_lm$coefficients[[1]],
        my_lm$coefficients[[1]] + my_lm$coefficients[[2]],
        my_lm$coefficients[[1]] + my_lm$coefficients[[3]])

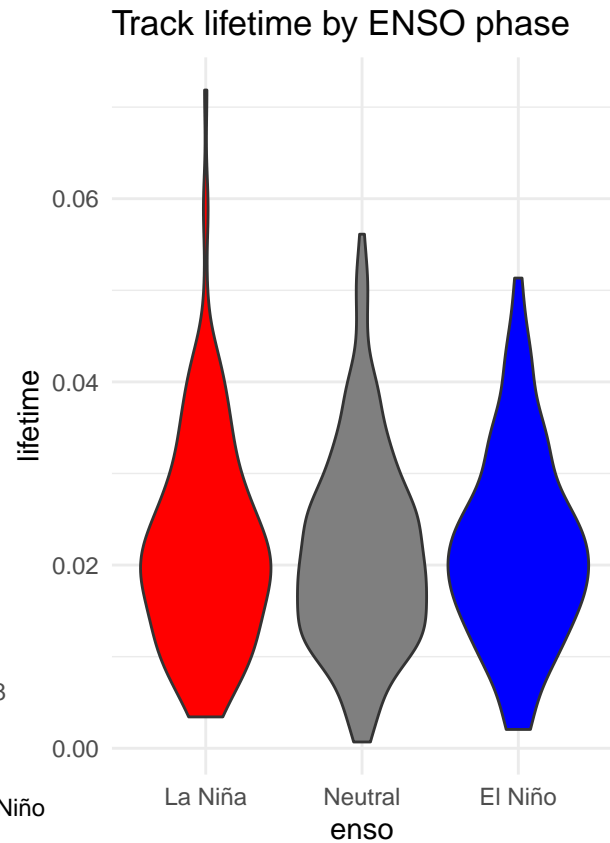
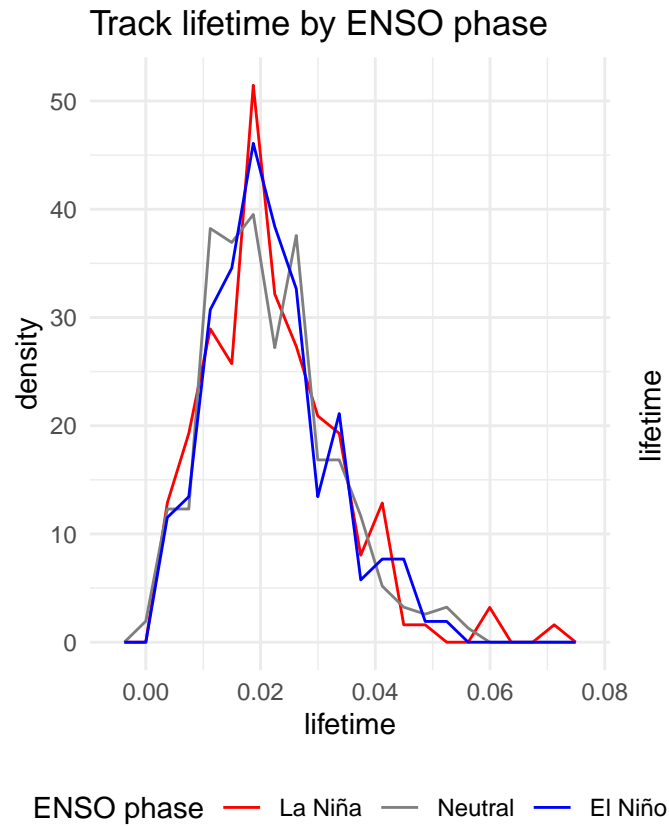
exp(mu)*365

## [1] 8.051463 7.839493 8.009802

p1 <- ggplot(data = tracks_1, aes(x = lifetime, y = after_stat(density))) +
  geom_freqpoly(aes(color= enso), bins=20) +
  theme_minimal() +
  theme(legend.position = 'bottom') +
  labs(title = "Track lifetime by ENSO phase") +
  scale_color_manual(
    name = "ENSO phase",
    labels = enso_labels,
    values = c("red", "grey50", "blue")
  )

p2 <- ggplot(data = tracks_1, aes(x = enso, y = lifetime)) +
  geom_violin(aes(fill = enso)) +
  theme_minimal() +
  theme(legend.position = "none") +
  labs(title = "Track lifetime by ENSO phase") +
  scale_fill_manual(
    name = "ENSO phase",
    labels = enso_labels,
    values = c("red", "grey50", "blue")
  ) +
  scale_x_discrete(labels = enso_labels)

grid.arrange(p1,p2, ncol = 2)
```



```
(cor_table <- cor(tracks_1 %>% select(max_sw, lat, lon, lifetime)))
```

```
##           max_sw      lat      lon  lifetime
## max_sw    1.0000000 0.1149604 0.16396426 0.50741888
## lat       0.1149604 1.0000000 0.27607626 0.21135135
## lon       0.1639643 0.2760763 1.00000000 0.01262047
## lifetime 0.5074189 0.2113514 0.01262047 1.00000000
```

```
corrplot(cor_table, method = "circle")
```

