TC data analysis – SI and SP basins

Melissa Renard

2025-02-19

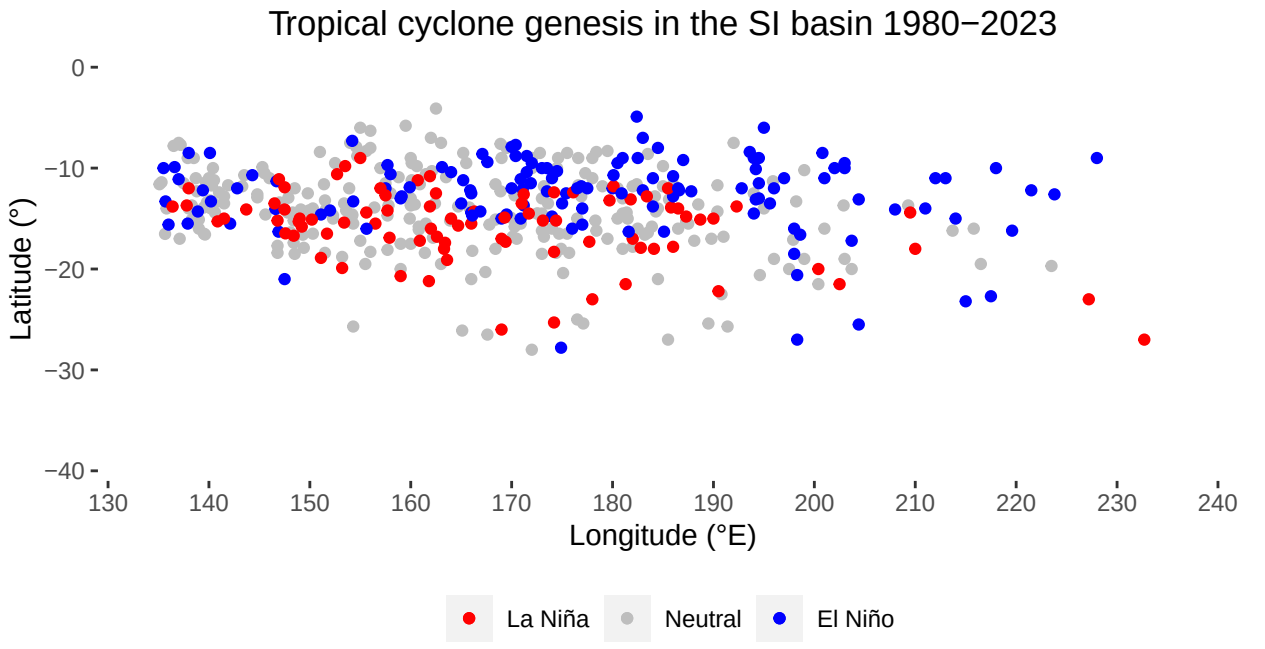
```
# CHOOSE BASIN HERE -- 'SI' OR 'SP'.
basin_choice <- 'SP'

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)
```

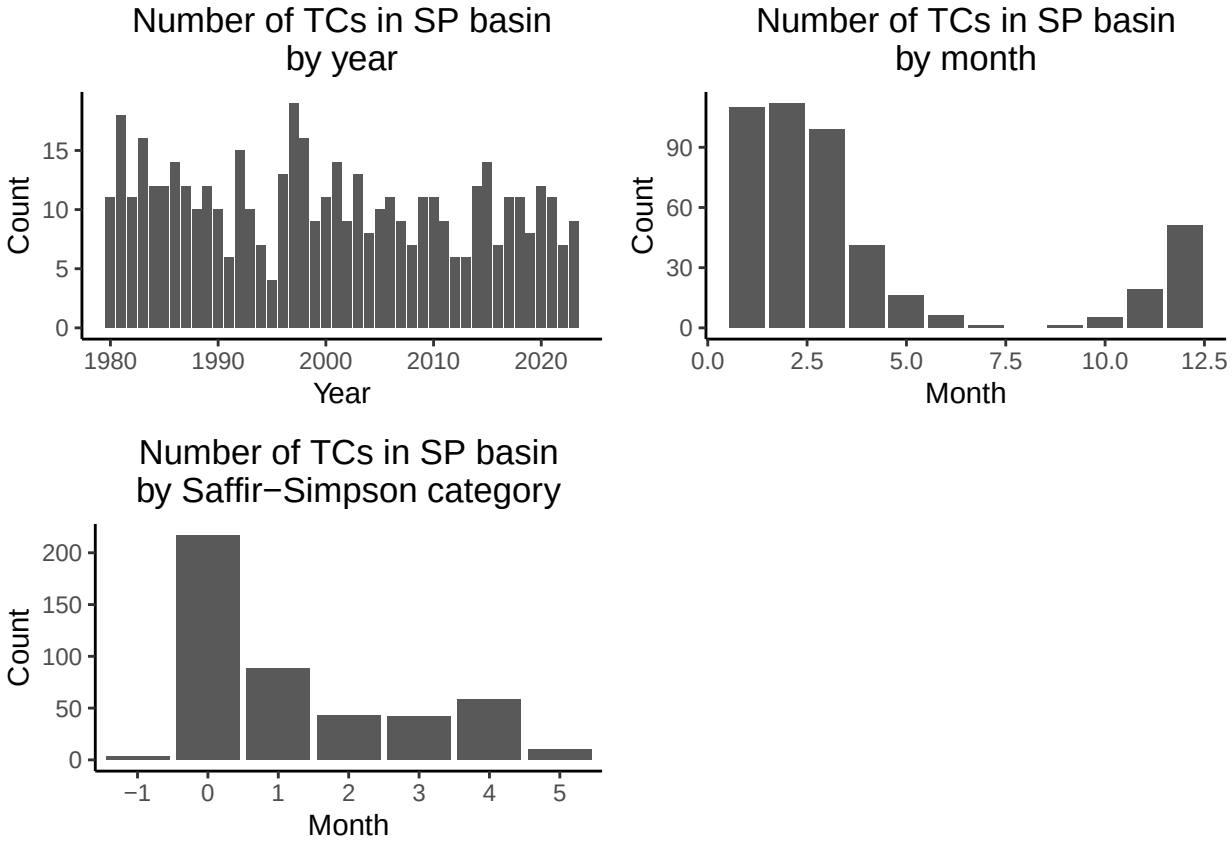
```r
tracks_1 <- tracks_1 %>% left_join(enso_df, by = join_by(enso_year==year))
```

```r
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
  )
```




```r
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
## -2.3568  -0.8705  -0.2227   0.3832   2.2775
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  2.37220    0.04657   50.93   <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: 45.365  on 42  degrees of freedom
## Residual deviance: 45.365  on 42  degrees of freedom
## AIC: 226.89
##
## Number of Fisher Scoring iterations: 4
```

```
mu <- my_lm$coefficients[[1]]
exp(mu)
```

```
## [1] 10.72093
```

```
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
## -2.4654  -0.7810  -0.1300   0.4789   1.8673
##
## Coefficients:
##                         Estimate Std. Error z value Pr(>|z|)
## (Intercept)               2.2336     0.1091  20.471   <2e-16 ***
## enso_df_TC$ensoNeutral    0.1098     0.1261   0.871   0.3839
## enso_df_TC$ensoNino       0.3080     0.1406   2.190   0.0285 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 45.365  on 42  degrees of freedom
## Residual deviance: 40.018  on 40  degrees of freedom
## AIC: 225.54
##
## Number of Fisher Scoring iterations: 4
```

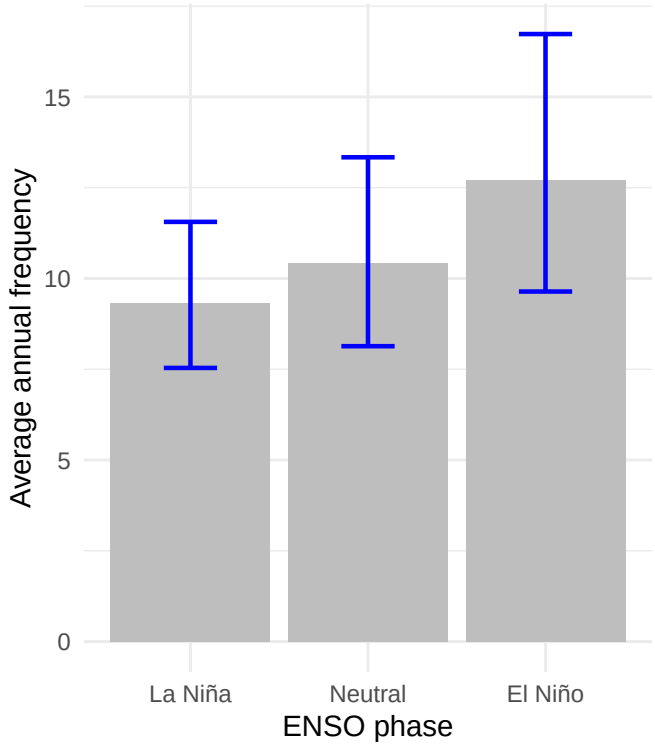
```
mu <- c(my_lm$coefficients[[1]],
        my_lm$coefficients[[1]] + my_lm$coefficients[[2]],
        my_lm$coefficients[[1]] + my_lm$coefficients[[3]])

exp(mu)
```

```
## [1]  9.333333 10.416667 12.700000
```

Estimated mean frequency of TCs in SI basin in La Niña, neutral and El Niño phases, respectively: 9.33, 10.42, 12.7

```
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.
##   135.1   153.2   169.0   169.1   182.5   232.7
```

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

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  -28.00  -16.10  -13.70  -14.01  -11.40   -4.10
```

```
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
## -41.743 -14.738  -2.296  14.204  64.762
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  167.938      2.187  76.795  < 2e-16 ***
## ensoNeutral   -2.642      2.528  -1.045  0.29642
## ensoNino       9.305      2.819   3.301  0.00104 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 20.04 on 458 degrees of freedom
## Multiple R-squared:  0.06196,    Adjusted R-squared:  0.05786
## F-statistic: 15.13 on 2 and 458 DF,  p-value: 4.355e-07
```

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

mu
```

```
## [1] 167.9379 165.2956 177.2427
```

Estimated mean longitude of TCs in La Niña, neutral and El Niño phases, respectively: 167.938, 165.296, 177.243.

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 = east170 ~ enso, family = binomial, data = tracks_1)
##
## Deviance Residuals:
##    Min      1Q  Median      3Q     Max
## -1.456  -1.037  -1.019   1.324   1.345
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -0.38566    0.22229  -1.735 0.082746 .
## ensoNeutral  0.04645    0.25666   0.181 0.856396
## ensoNino     1.02031    0.29015   3.516 0.000437 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 638.30  on 460  degrees of freedom
## Residual deviance: 616.75  on 458  degrees of freedom
## AIC: 622.75
##
## 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] 0.4047619 0.4160000 0.6535433
```

Estimated probability of a TC produced east of 75°E in La Niña, neutral and El Niño phases, respectively: 0.4047619, 0.416, 0.6535433

Intensity

Intensity based on ENSO phases

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

summary(my_lm)
```

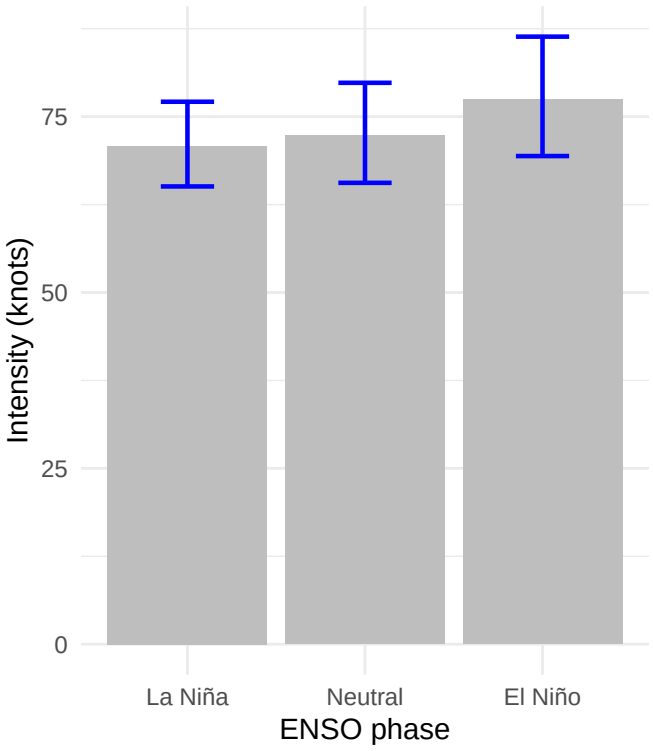
```
##
## Call:
## glm(formula = max_sw ~ enso, family = Gamma(link = log), data = tracks_1)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.0460  -0.3281  -0.1244   0.2374   0.9083
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  4.26044    0.04331  98.370   <2e-16 ***
## ensoNeutral  0.02110    0.05006   0.422    0.674
## ensoNino     0.08880    0.05583   1.591    0.112
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## (Dispersion parameter for Gamma family taken to be 0.1575672)
##
##     Null deviance: 70.847  on 460  degrees of freedom
## Residual deviance: 70.324  on 458  degrees of freedom
## AIC: 4352.1
##
## 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)
```

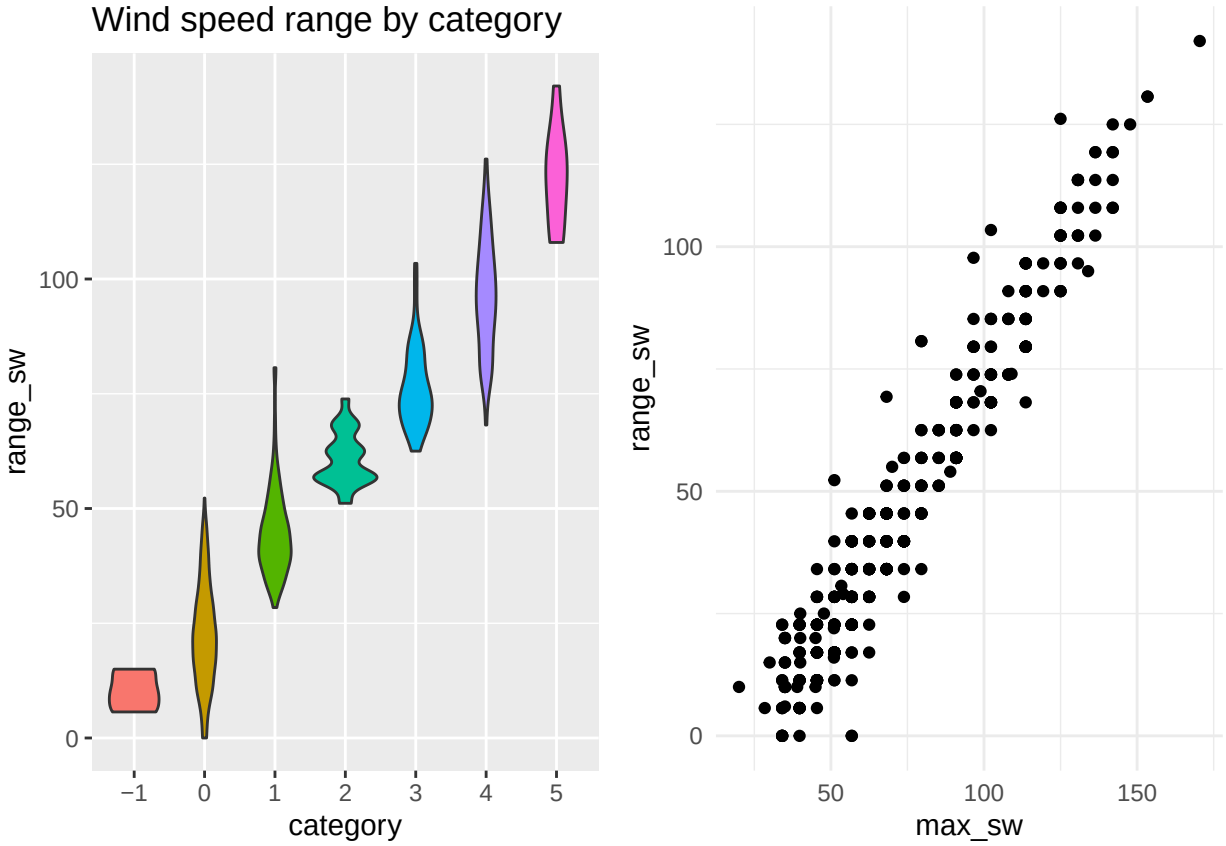
```
## [1] 70.84091 72.35182 77.41947
```

```
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)
```

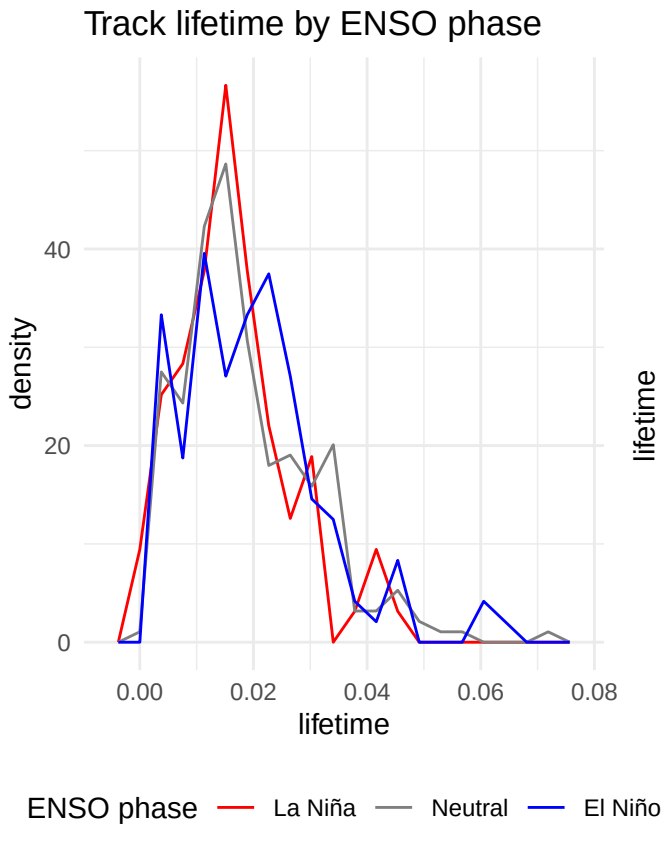
```
##
## Call:
## glm(formula = lifetime ~ enso, family = Gamma(link = log), data = tracks_1)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.1137  -0.5070  -0.1001   0.3158   1.7405
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -4.09409    0.06634 -61.717   <2e-16 ***
## ensoNeutral  0.11815    0.07668   1.541   0.1240
## ensoNino     0.16217    0.08551   1.897   0.0585 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for Gamma family taken to be 0.3696507)
##
##     Null deviance: 198.76  on 460  degrees of freedom
## Residual deviance: 197.41  on 458  degrees of freedom
## AIC: -2913.9
```

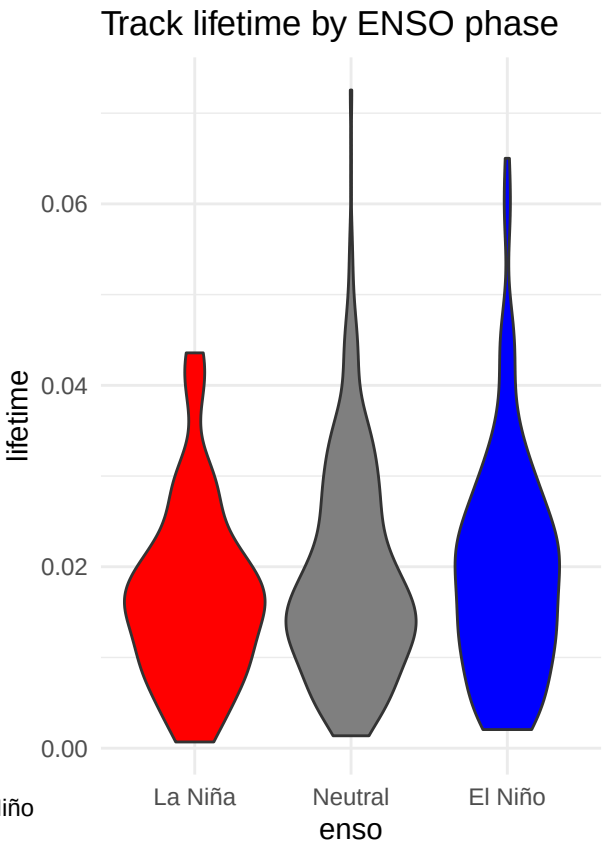
```
##
## Number of Fisher Scoring iterations: 5
```

```
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] 6.084870 6.847976 7.156188
```

```
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)
```

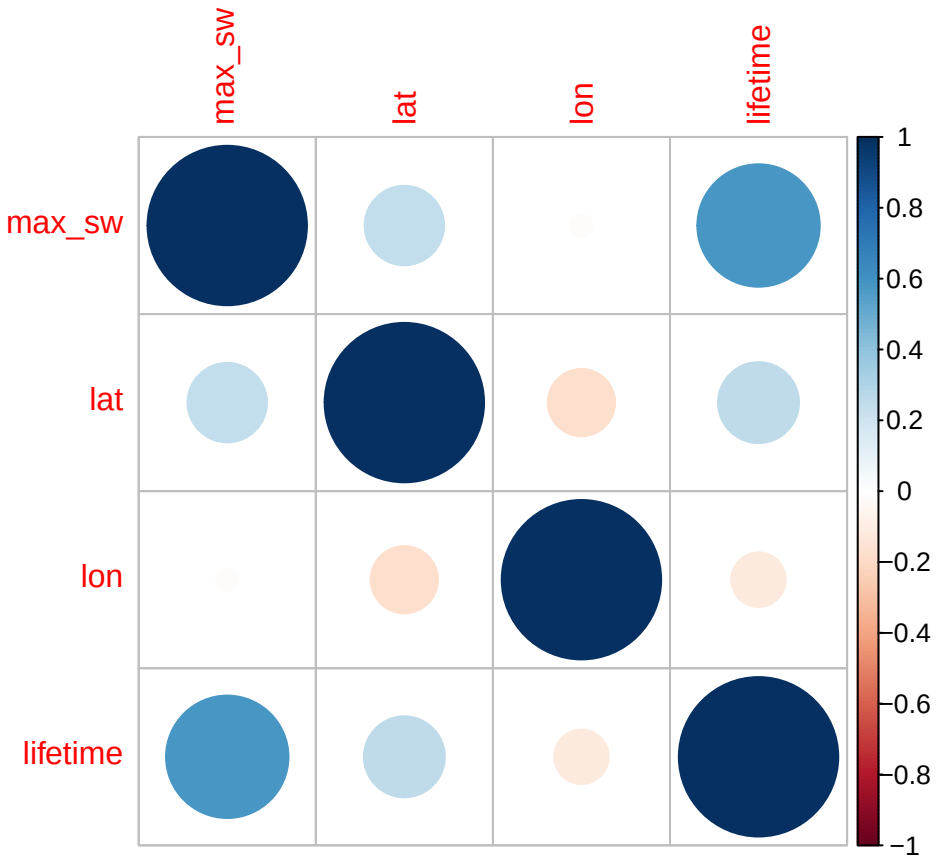





```
(cor_table <- cor(tracks_1 %>% select(max_sw, lat, lon, lifetime)))
```

```
##                max_sw        lat         lon   lifetime
## max_sw     1.00000000  0.2480966 -0.01874712  0.5895773
## lat        0.24809658  1.0000000 -0.17821279  0.2578512
## lon       -0.01874712 -0.1782128  1.00000000 -0.1170194
## lifetime   0.58957730  0.2578512 -0.11701936  1.0000000
```

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


max_sw
lat
lon
lifetime
max_sw
lat
lon
lifetime
1
0.8
0.6
0.4
0.2
0
−0.2
−0.4
−0.6
−0.8
−1