Elaborazioni Gambero2025

Gianandrea LP

2025-06-25

```
    # -------------------------------------
    # dir.create('input', showWarnings = FALSE)
    # dir.create('output', showWarnings = FALSE)
    # dir.create('plots', showWarnings = FALSE)
    # dir.create('R-script', showWarnings = FALSE)
    library(pacman)
    p_load(tidyverse,stargazer,skimr,janitor,lubridate,
        ggpubr,rstatix,openxlsx, patchwork, knitr, zoo)

    options(scipen = 9999)
    theme_set(theme_minimal(base_size = 18))
    # -------------------------------------
```

Gambero 2025

Serie storica triennale su *Procambarus clarkii* del lago Trasimeno. Anni di riferimento: luglio 2018 - maggio 2021

```
# import data
pc <- read.xlsx('input/Gamberi morfometria 3 anni_.xlsx',
                detectDates = T)
glimpse(pc)
```

```
Rows: 3,737
Columns: 13
$ data            <date> 2018-07-04, 2018-07-04, 2018-07-04, 2018-07-04, 2018-~
$ n               <dbl> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16,~
$ sesso           <chr> "M", "M", "M", "M", "M", "M", "M", "M", "M", "M", "M",~
$ lt.carapace     <dbl> 43.28, 54.11, 52.73, 81.55, 72.00, 52.21, 54.89, 59.21~
```

```
$ peso.tot         <dbl> 16.90, 34.50, 30.20, 35.40, 28.50, 33.60, 37.00, 39.10~
$ peso.addome      <dbl> 5.0, 8.2, 7.0, 8.2, 6.0, 7.5, 7.7, 7.8, 8.2, 6.0, 6.4,~
$ peso.epato       <dbl> 0.7, 1.7, 1.8, 1.9, 0.9, 2.1, 1.3, 1.7, 1.6, 1.4, 1.1,~
$ Hiw              <dbl> 4.142012, 4.927536, 5.960265, 5.367232, 3.157895, 6.25~
$ TwB              <dbl> 29.58580, 23.76812, 23.17881, 23.16384, 21.05263, 22.3~
$ SA               <chr> "SI", "SI", "SI", "SI", "SI", "SI", "SI", "SI", "SI", ~
$ muta             <chr> "NO", "NO", "NO", "NO", "NO", "NO", "NO", "NO", "NO", ~
$ maturità.gonadi <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA~
$ peso.gonadi      <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA~
```

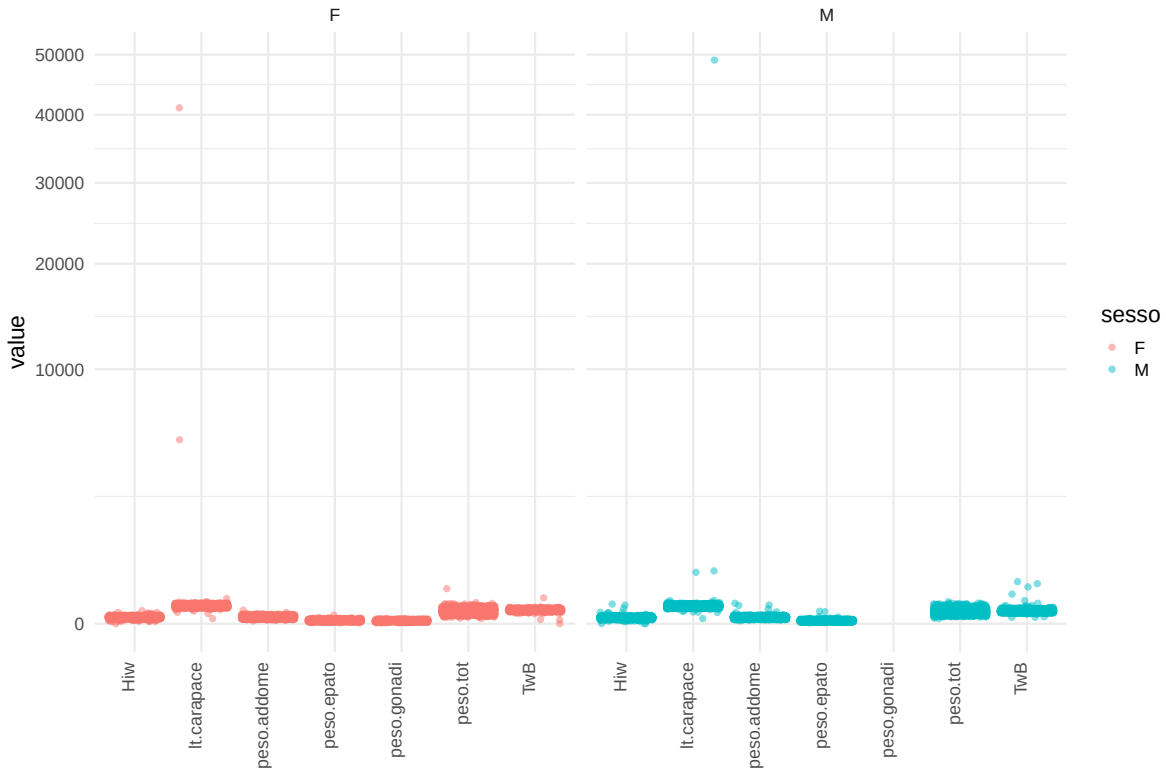
controllo qualità dei dati

```
pc %>%
  select(where(is.numeric),sesso,-n) %>%
  group_by(sesso) %>%
  get_summary_stats(type = 'five_number') %>%
  kable(digits = 2)
```

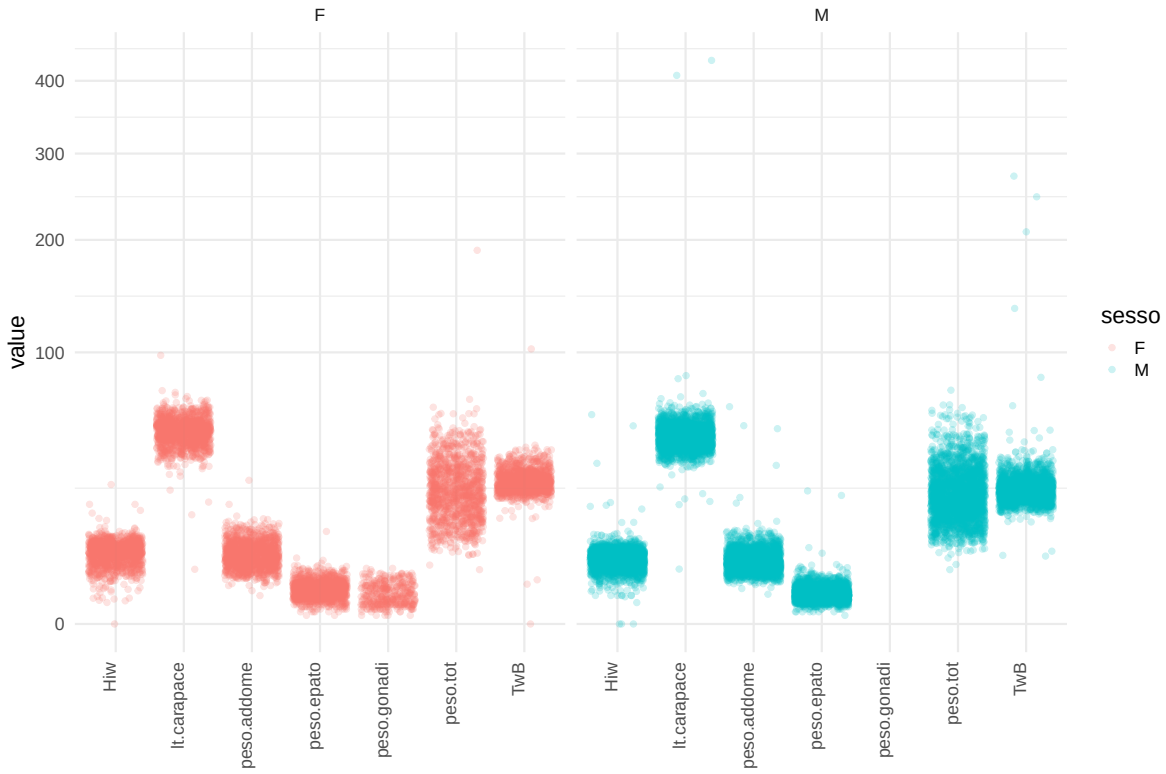
sesso	variable	n	min	max	q1	median	q3
F	lt.carapace	1243	4.06	41081.00	45.41	50.04	54.83
F	peso.tot	1243	4.00	189.30	16.85	23.50	31.50
F	peso.addome	1242	1.09	28.00	4.80	6.30	8.30
F	peso.epato	1242	0.10	11.60	1.10	1.50	2.20
F	Hiw	1243	0.00	26.36	5.38	7.01	8.33
F	TwB	1243	0.00	102.56	24.91	27.16	30.02
F	peso.gonadi	409	0.10	4.20	0.60	1.10	2.00
M	lt.carapace	2494	4.09	49039.00	43.63	47.41	51.45
M	peso.tot	2493	4.00	74.10	16.80	21.90	28.70
M	peso.addome	2494	1.20	64.00	4.30	5.30	6.56
M	peso.epato	2491	0.10	24.00	0.90	1.20	1.60
M	Hiw	2493	0.00	59.42	4.65	5.75	6.74
M	TwB	2493	6.25	271.94	21.65	24.10	26.67

```
pc %>%
  select(where(is.numeric),sesso) %>%
  pivot_longer(lt.carapace:peso.gonadi,
               names_to = 'var', values_to = 'value') %>%
  ggplot(aes(var,value))+
  geom_jitter(aes(col=sesso),alpha=0.5, size=2)+
  scale_y_sqrt()+
```

```
  facet_grid(~sesso)+
  theme(axis.text.x = element_text(angle = 90, vjust = 0, hjust=1))+
  labs(x='')
```




```
pc %>%
  select(where(is.numeric),sesso) %>%
  pivot_longer(lt.carapace:peso.gonadi,
               names_to = 'var', values_to = 'value') %>%
  filter(value<800) %>%
  ggplot(aes(var,value))+
  geom_jitter(aes(col=sesso),alpha=0.2, size=2)+
  scale_y_sqrt()+
  facet_grid(~sesso)+
  theme(axis.text.x = element_text(angle = 90, vjust = 0, hjust=1))+
  labs(x='')
```




```
# add month and year
pc <- pc %>%
  mutate(month=month(data),
         year=year(data),
         date=floor_date(data, unit = 'month'))
```

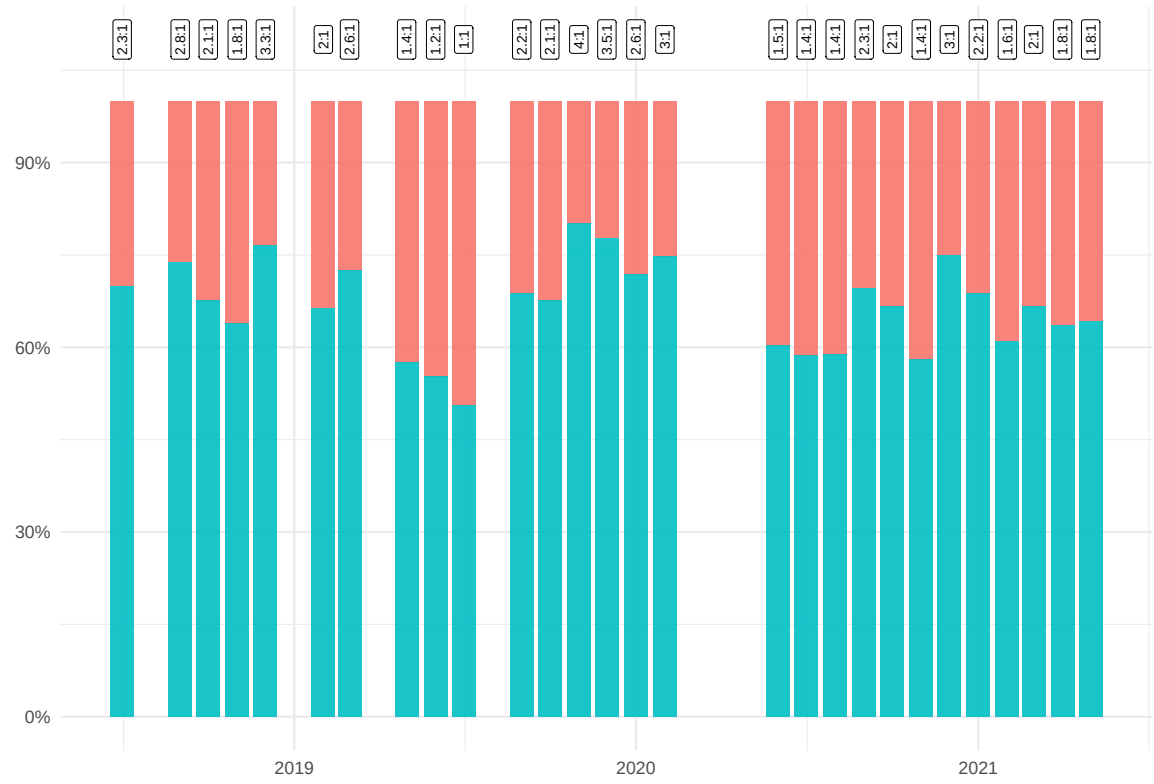
```
clark.sex <- pc %>%
  group_by(date,sesso) %>%
  summarise(n=n()) %>%
  mutate(tot=cumsum(n))
```

```
`summarise()` has grouped output by 'date'. You can override using the
`.groups` argument.
```

```
clark.sexratioF <- clark.sex %>%
  filter(sesso=='F')
clark.sexratioM <- clark.sex %>%
  filter(sesso=='M')
```

```r
sexratio <- cbind(data.frame(ratio=round(clark.sexratioM$n/clark.sexratioF$n,1)),
                  distinct(pc,date))
```

```r
pc %>%
  count(sesso,date) %>%
  group_by(date) %>%
  mutate(perc=n/sum(n)) %>%
  ggplot(aes(date,perc,group = sesso))+
  geom_bar(stat = 'identity', alpha=0.9, aes(fill=sesso))+
  scale_y_continuous(labels = scales::percent_format())+
  labs(x='',y='')+
  geom_label(data=sexratio,
             aes(x=date,y=1.1,
             label=paste0(ratio,':1')),
             inherit.aes = FALSE,
             angle=90)+
  theme(legend.position = 'none')
```

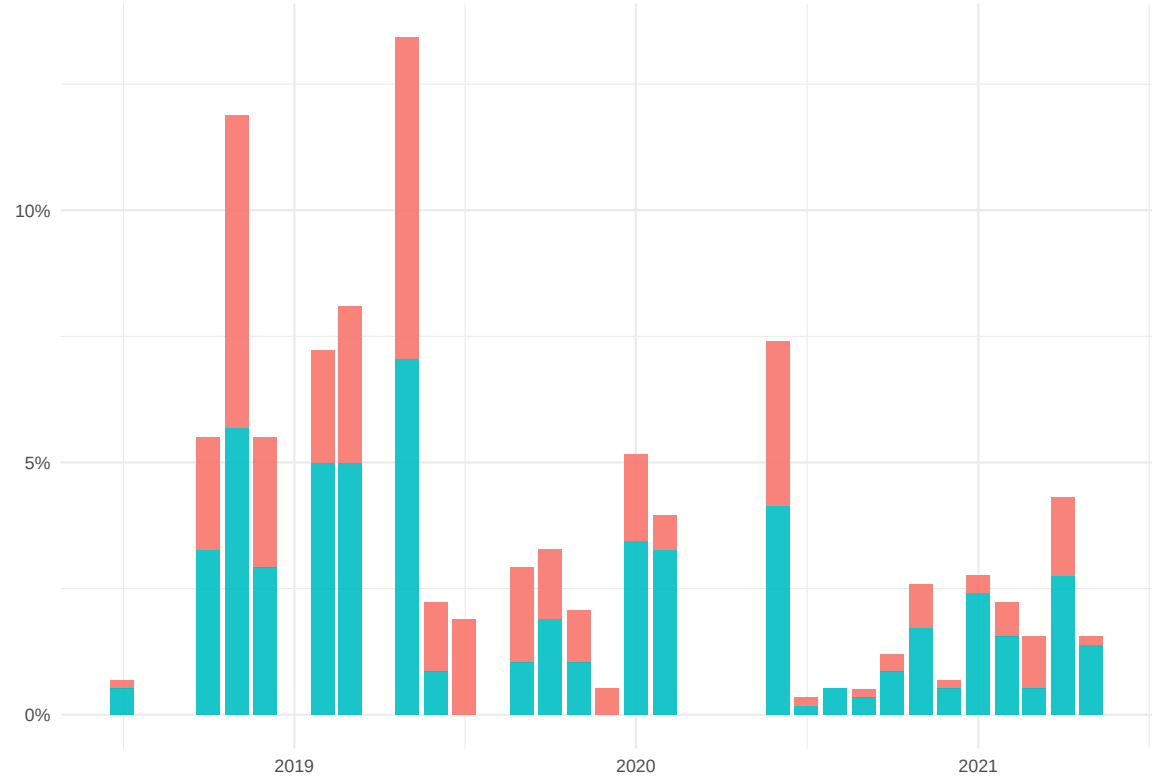


```r
mute <- pc %>%
  filter(muta=='SI') %>%
  count(sesso,date) %>%
  mutate(perc=n/sum(n))

mute %>%
  mutate(perc=round(perc*100,2)) %>%
  pivot_wider(id_cols=-n,names_from = sesso,values_from = perc) %>%
  kable()
```

date	F	M
2018-07-01	0.17	0.52
2018-10-01	2.24	3.27
2018-11-01	6.20	5.68
2018-12-01	2.58	2.93
2019-02-01	2.24	4.99
2019-03-01	3.10	4.99
2019-05-01	6.37	7.06
2019-06-01	1.38	0.86
2019-07-01	1.89	NA
2019-09-01	1.89	1.03
2019-10-01	1.38	1.89
2019-11-01	1.03	1.03
2019-12-01	0.52	NA
2020-01-01	1.72	3.44
2020-02-01	0.69	3.27
2020-06-01	3.27	4.13
2020-07-01	0.17	0.17
2020-09-01	0.17	0.34
2020-10-01	0.34	0.86
2020-11-01	0.86	1.72
2020-12-01	0.17	0.52
2021-01-01	0.34	2.41
2021-02-01	0.69	1.55
2021-03-01	1.03	0.52
2021-04-01	1.55	2.75
2021-05-01	0.17	1.38
2020-08-01	NA	0.52

```
mute %>% ggplot(aes(date,perc,group = sesso))+
  geom_bar(stat = 'identity', alpha=0.9, aes(fill=sesso))+
  scale_y_continuous(labels = scales::percent_format())+
  labs(x='',y='')+
  theme(legend.position = 'none')
```



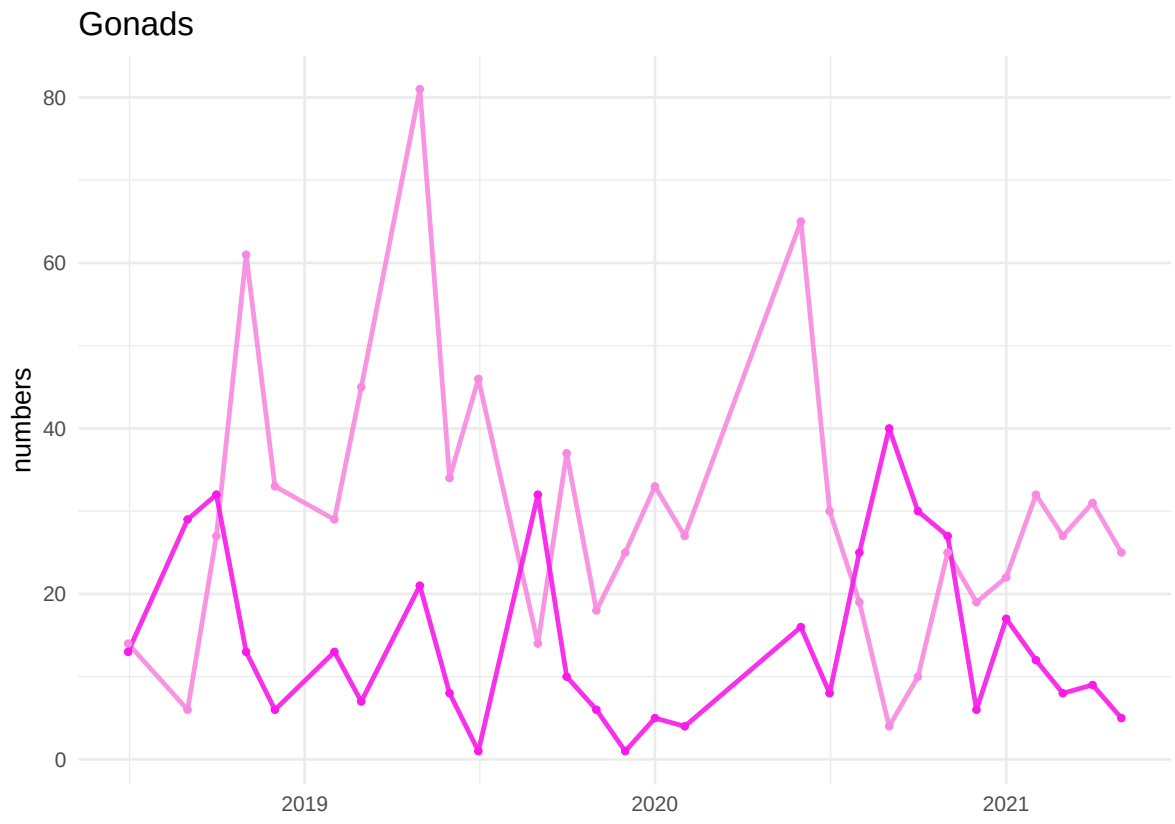

```
unique(pc$maturità.gonadi)
```

```
 [1] NA                      "MATURE"                "IMMATURE"
 [4] "MATURE "               "CON LARVE"             "ASSENTI"
 [7] "IN VIA DI MATURAZIONE" "UOVA PLEOPODALI"       "DEGENERATE"
[10] "INIZIO MATURAZIONE"    "QUASI MATURE"
```

```
# corregge i nomi
pc$maturità.gonadi <- as.factor(pc$maturità.gonadi)
levels(pc$maturità.gonadi)[8] <- 'MATURE'
```

```r
# crea una nuova variabile con stadio gonadi: 0, 1
pc.f <- pc %>%
  filter(sesso=='F') %>%
  mutate(gonadBin=ifelse(maturità.gonadi=='MATURE','1','0'))

pc.f %>%
  count(date,gonadBin) %>%
  ggplot(aes(x=date,y=n, group=gonadBin,col=gonadBin)) +
  geom_line(alpha=0.9, lwd=1.5) +
  geom_point(size=2) +
  ggtitle('Gonads') + ylab('numbers')+ xlab('')+
  scale_color_manual(values = c('#f78ae0','#f71ae8'))+
  theme(legend.position = 'none')
```




```r
pc.f %>%
  count(date,gonadBin) %>%
  group_by(date) %>%
```

```
  mutate(perc=round(n/sum(n)*100,2)) %>%
  pivot_wider(id_cols = date,
              values_from = perc,
              names_from = gonadBin) %>%
  kable()
```

date	0	1
2018-07-01	51.85	48.15
2018-09-01	17.14	82.86
2018-10-01	45.76	54.24
2018-11-01	82.43	17.57
2018-12-01	84.62	15.38
2019-02-01	69.05	30.95
2019-03-01	86.54	13.46
2019-05-01	79.41	20.59
2019-06-01	80.95	19.05
2019-07-01	97.87	2.13
2019-09-01	30.43	69.57
2019-10-01	78.72	21.28
2019-11-01	75.00	25.00
2019-12-01	96.15	3.85
2020-01-01	86.84	13.16
2020-02-01	87.10	12.90
2020-06-01	80.25	19.75
2020-07-01	78.95	21.05
2020-08-01	43.18	56.82
2020-09-01	9.09	90.91
2020-10-01	25.00	75.00
2020-11-01	48.08	51.92
2020-12-01	76.00	24.00
2021-01-01	56.41	43.59
2021-02-01	72.73	27.27
2021-03-01	77.14	22.86
2021-04-01	77.50	22.50
2021-05-01	83.33	16.67

```
pc.f %>%
  count(date,gonadBin) %>%
  group_by(date) %>%
  mutate(perc=n/sum(n)) %>%
```

```r
ggplot(aes(date,perc,group = gonadBin, fill=gonadBin))+
geom_bar(stat = 'identity', alpha=0.9)+
scale_y_continuous(labels = scales::percent_format())+
labs(x='',y='')+
scale_fill_manual(values = c('#f78ae0','#f71ae8'))+
theme(legend.position = 'none')
```