STAT 426 Project 2

Tommy Ryan

12/12/2020

```
library(MASS)
library(ggplot2)
library(dplyr)
library(ggpubr)
library(jtools)
library(lme4)
```

```
data("bacteria")
help("bacteria")
```

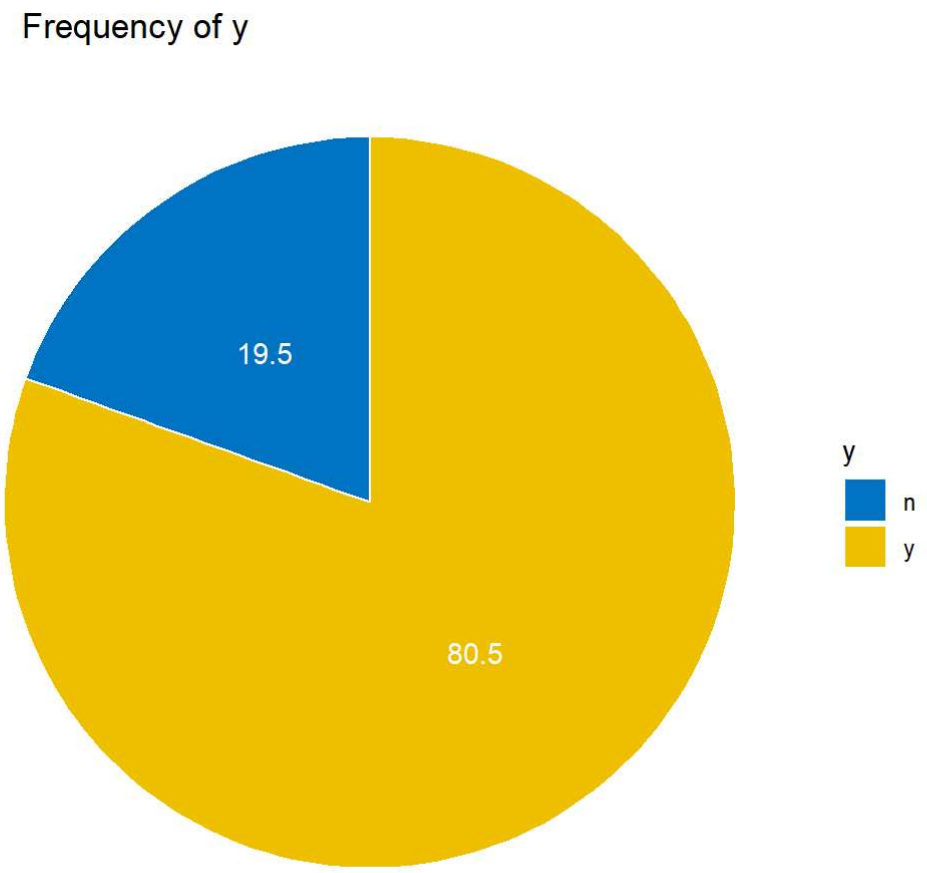
```
## starting httpd help server ... done
```

```
df1 <- bacteria %>%
  group_by(y) %>%
  summarise(counts = n())
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```
df1 <- df1 %>%
  arrange(desc(y)) %>%
  mutate(prop = round(counts*100/sum(counts), 1),
         lab.ypos = cumsum(prop) - 0.5*prop)
```

```
p1 <- ggplot(df1, aes(x = "", y = prop, fill = y)) +
  geom_bar(width = 1, stat = "identity", color = "white") +
  geom_text(aes(y = lab.ypos, label = prop), color = "white") +
  coord_polar("y", start = 0) +
  fill_palette("jco")+
  theme_void()

p1 + ggtitle("           Frequency of y")
```

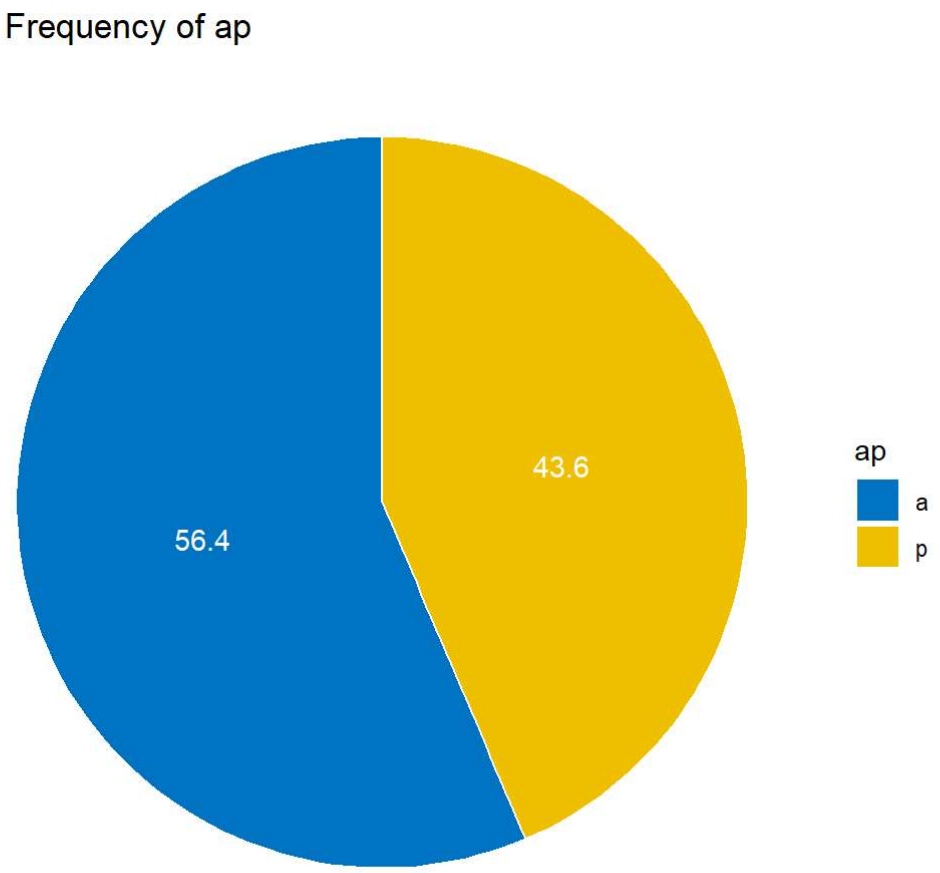



```
df2 <- bacteria %>%
  group_by(ap) %>%
  summarise(counts = n())
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```
df2 <- df2 %>%
  arrange(desc(ap)) %>%
  mutate(prop = round(counts*100/sum(counts), 1),
         lab.ypos = cumsum(prop) - 0.5*prop)
```

```
p2 <- ggplot(df2, aes(x = "", y = prop, fill = ap)) +
  geom_bar(width = 1, stat = "identity", color = "white") +
  geom_text(aes(y = lab.ypos, label = prop), color = "white") +
  coord_polar("y", start = 0) +
  fill_palette("jco") +
  theme_void()

p2 + ggtitle( "        Frequency of ap")
```

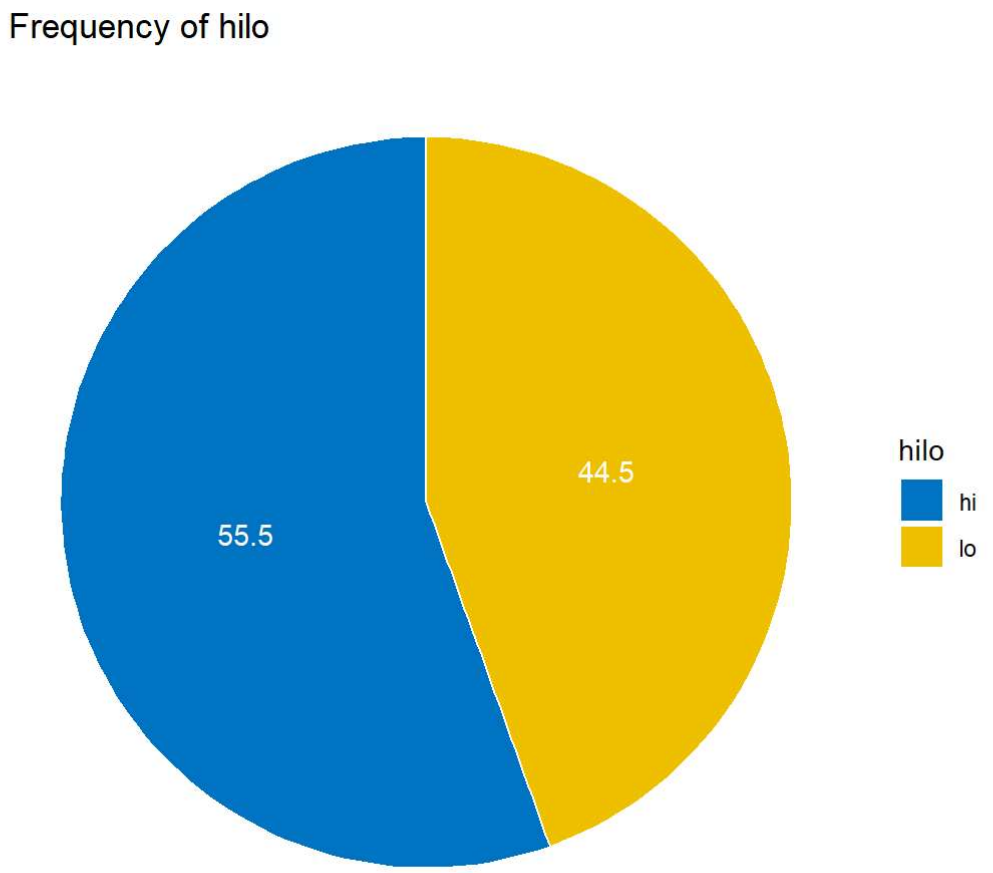



```
df3 <- bacteria %>%
  group_by(hilo) %>%
  summarise(counts = n())
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```
df3 <- df3 %>%
  arrange(desc(hilo)) %>%
  mutate(prop = round(counts*100/sum(counts), 1),
         lab.ypos = cumsum(prop) - 0.5*prop)
```

```
p3 <- ggplot(df3, aes(x = "", y = prop, fill = hilo)) +
  geom_bar(width = 1, stat = "identity", color = "white") +
  geom_text(aes(y = lab.ypos, label = prop), color = "white") +
  coord_polar("y", start = 0) +
  fill_palette("jco") +
  theme_void()

p3 + ggtitle( "    Frequency of hilo")
```



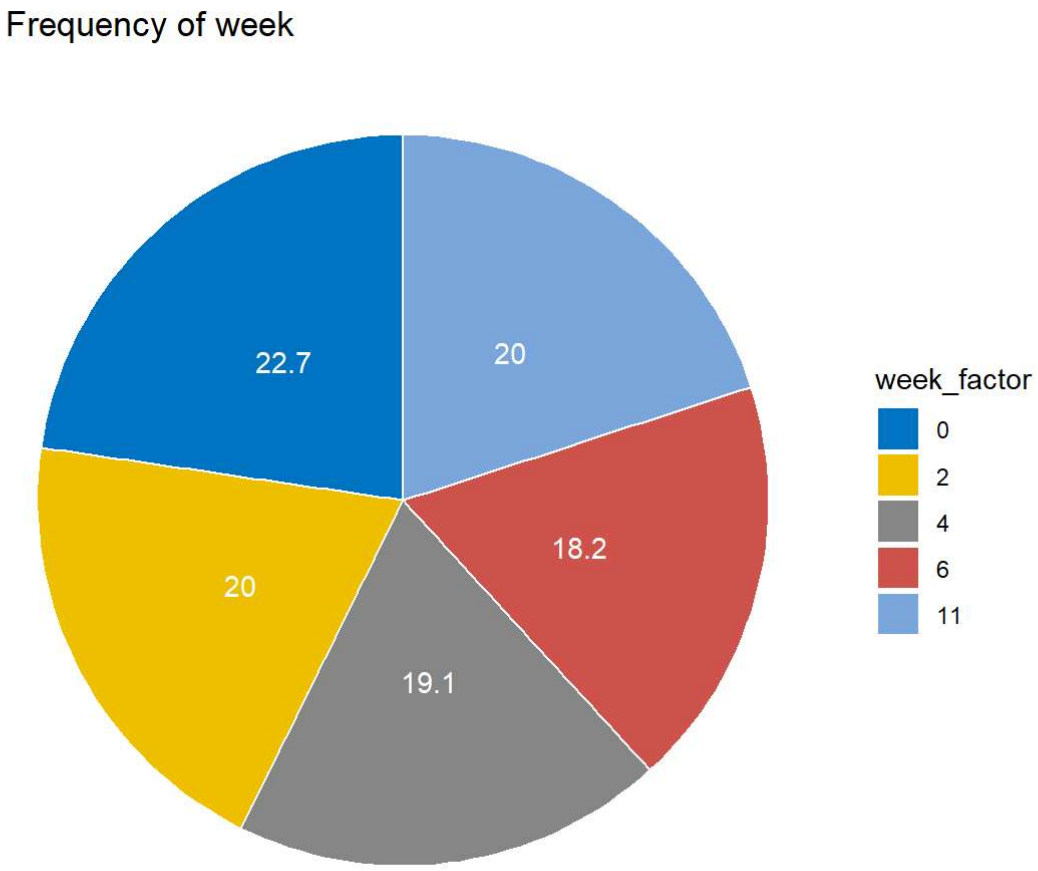

```
bacteria$week_factor = as.factor(bacteria$week)
```

```
df4 <- bacteria %>%
  group_by(week_factor) %>%
  summarise(counts = n())
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

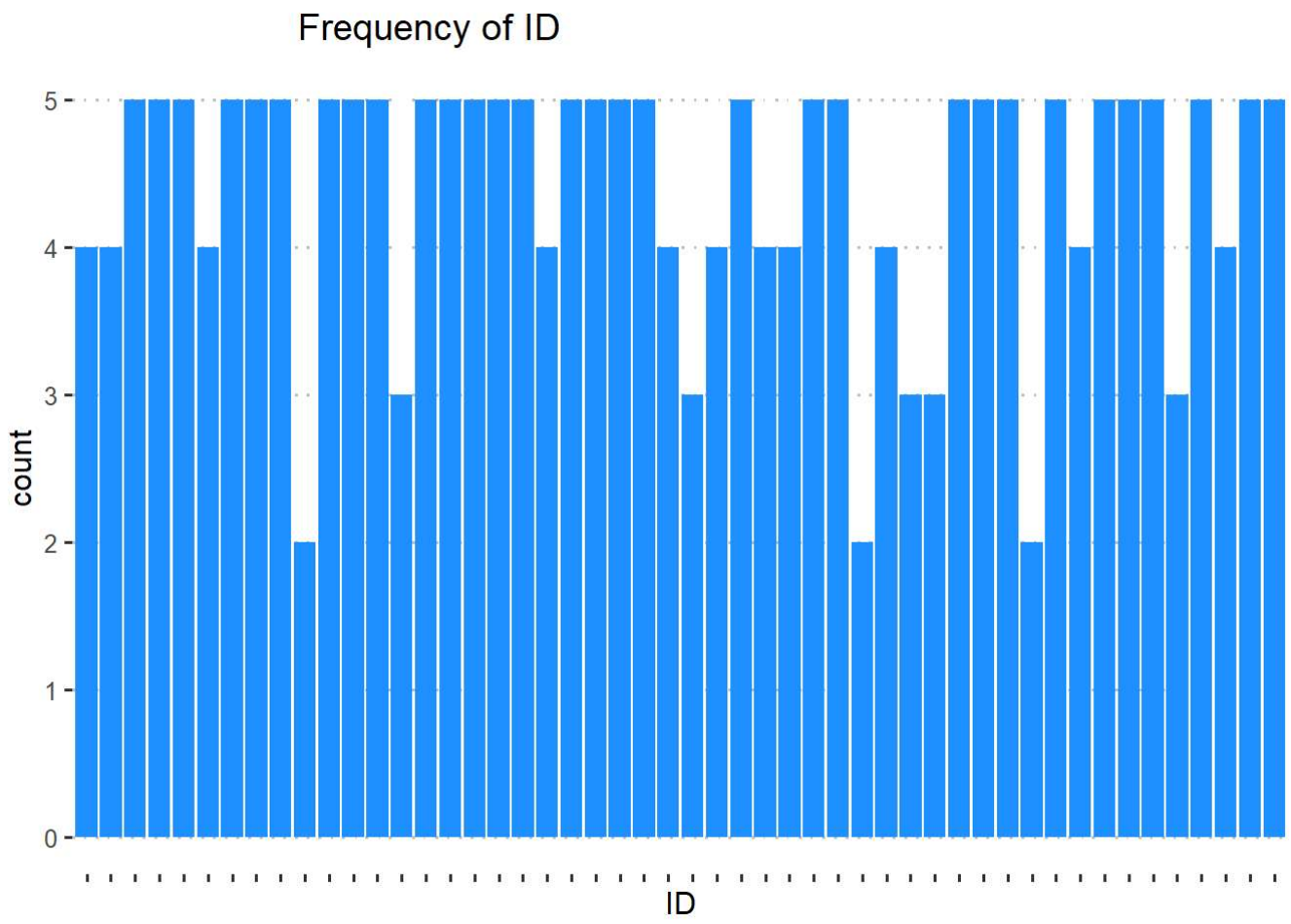
```
df4 <- df4 %>%
  arrange(desc(week_factor)) %>%
  mutate(prop = round(counts*100/sum(counts), 1),
         lab.ypos = cumsum(prop) - 0.5*prop)
```

```
p4 <- ggplot(df4, aes(x = "", y = prop, fill = week_factor)) +
  geom_bar(width = 1, stat = "identity", color = "white") +
  geom_text(aes(y = lab.ypos, label = prop), color = "white") +
  coord_polar("y", start = 0) +
  fill_palette("jco") +
  theme_void()

p4 + ggtitle( "      Frequency of week")
```




```
data("bacteria")
```

```
p5 <- ggplot(bacteria, aes(ID)) +
  geom_bar(fill = "dodgerblue") +
  theme_pubclean()
p5 + ggtitle( "                    Frequency of ID") +
  theme(axis.text.x = element_blank())
```

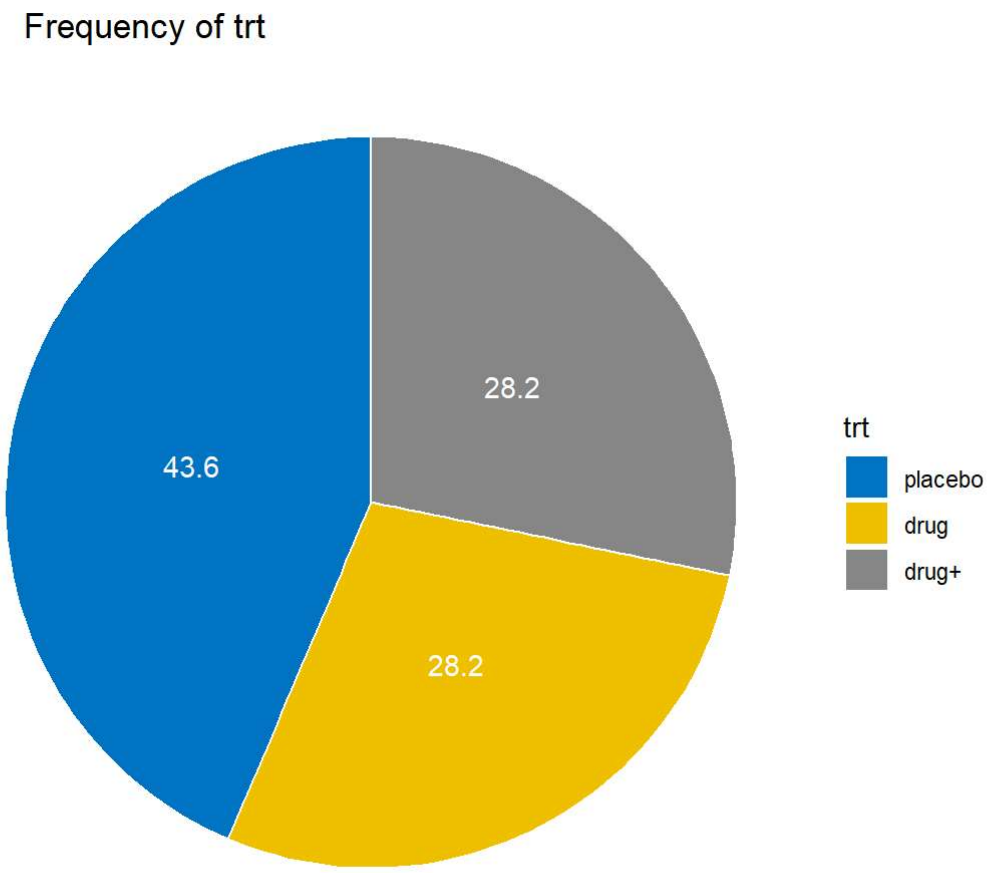



```
df6 <- bacteria %>%
  group_by(trt) %>%
  summarise(counts = n())
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```
df6 <- df6 %>%
  arrange(desc(trt)) %>%
  mutate(prop = round(counts*100/sum(counts), 1),
         lab.ypos = cumsum(prop) - 0.5*prop)
```

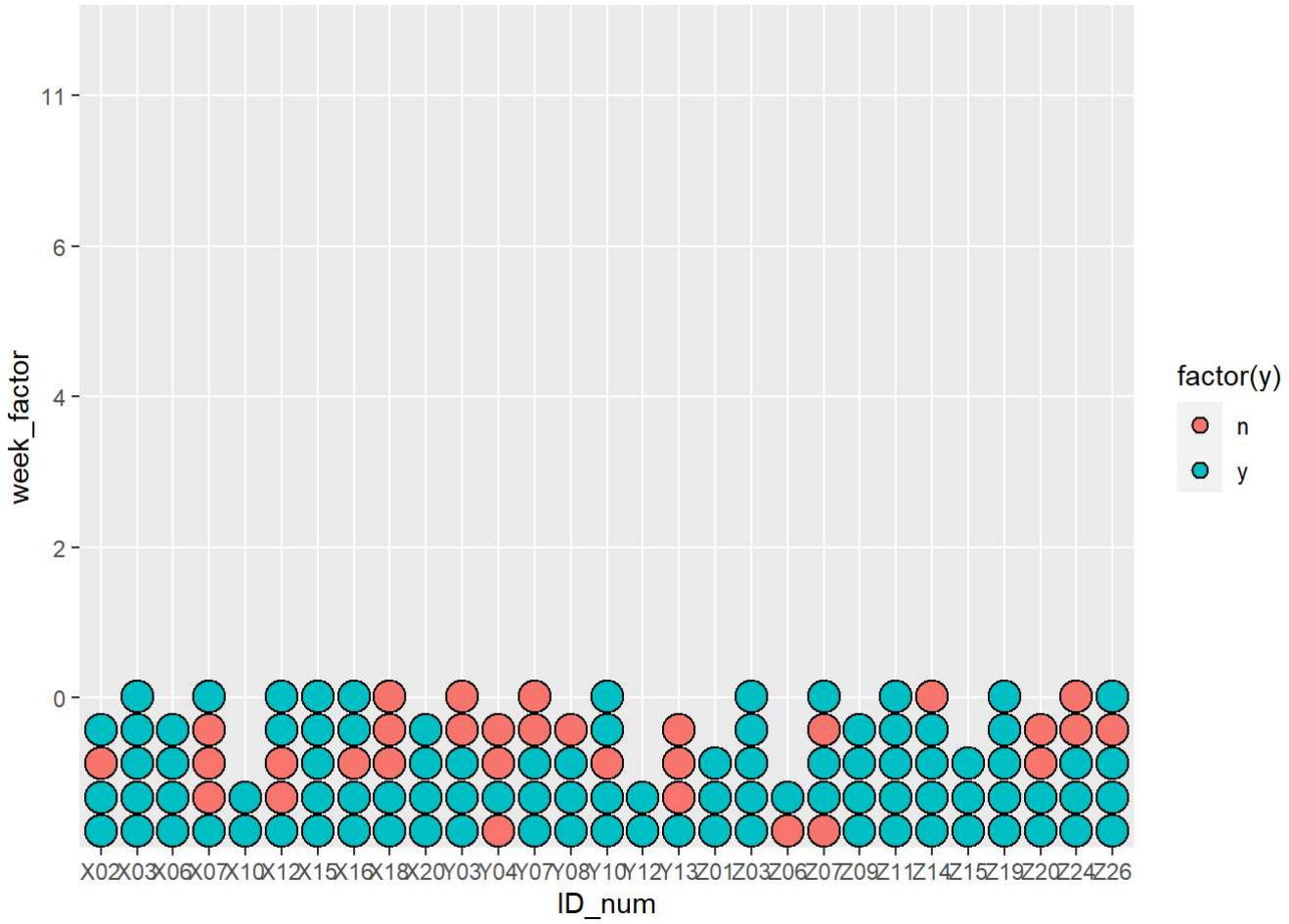
```
p6 <- ggplot(df6, aes(x = "", y = prop, fill = trt)) +
  geom_bar(width = 1, stat = "identity", color = "white") +
  geom_text(aes(y = lab.ypos, label = prop), color = "white") +
  coord_polar("y", start = 0) +
  fill_palette("jco") +
  theme_void()

p6 + ggtitle( "           Frequency of trt")
```




```
active_data <- bacteria[ which(bacteria$ap == 'a'), ]

active_data$week_factor = as.factor(active_data$week)

active_data$ID_num = as.character(active_data$ID)

placebo_data <- bacteria[ which(bacteria$ap == 'p'), ]

placebo_data$week_factor = as.factor(placebo_data$week)

placebo_data$ID_num = as.character(placebo_data$ID)
```

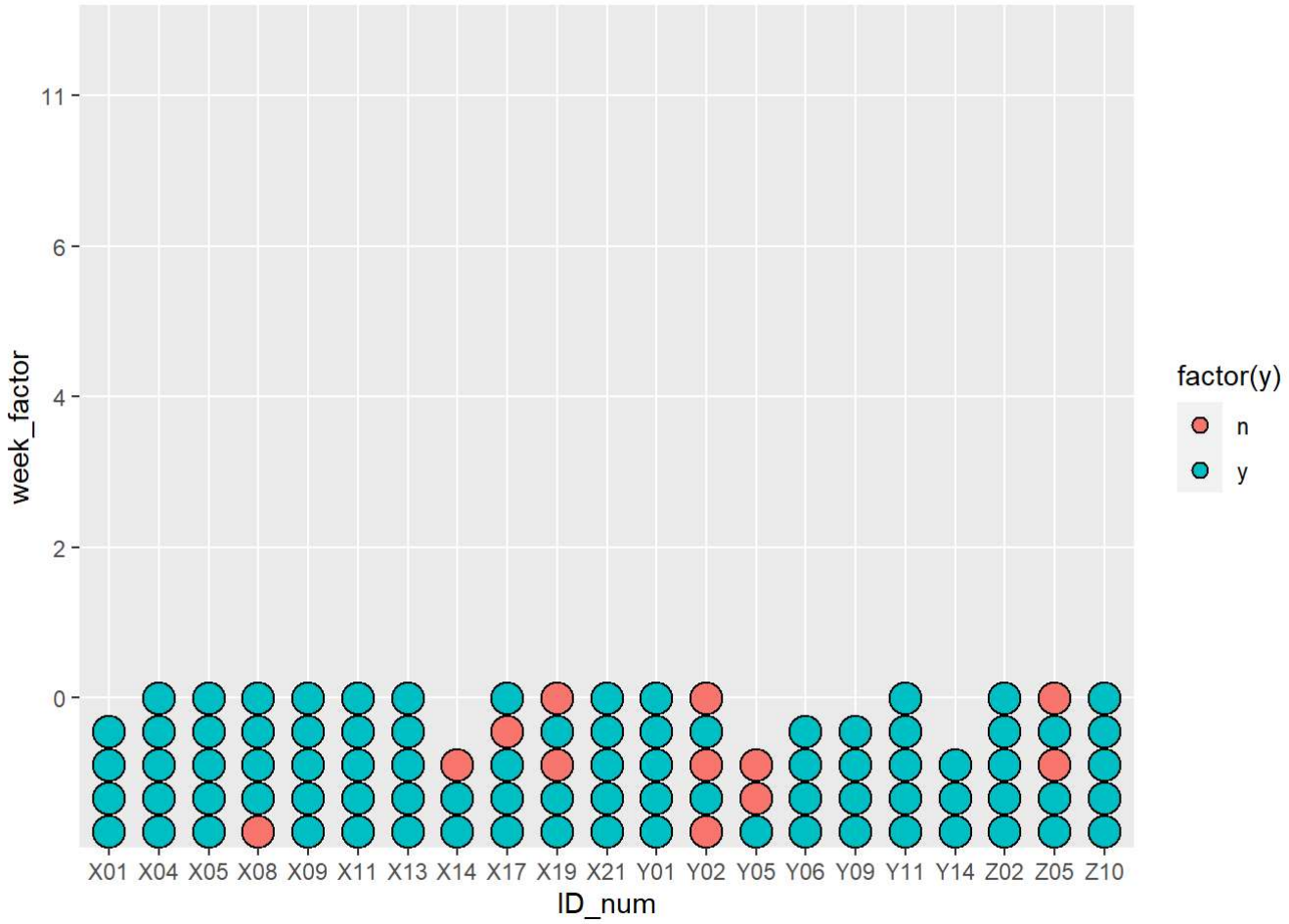
```
active_plot <- ggplot(active_data, aes(x = ID_num, y = week_factor, fill = factor(y))) +
  geom_dotplot(stackgroups = TRUE, stackdir = "up", binaxis = "x", binpositions = "all")
active_plot
```

```
## `stat_bindot()` using `bins = 30`. Pick better value with `binwidth`.
```




```
placebo_plot <- ggplot(placebo_data, aes(x = ID_num, y = week_factor, fill = factor(y))) +
  geom_dotplot(stackgroups = TRUE, stackdir = "up", binaxis = "x", binpositions = "all")
placebo_plot
```

```
## `stat_bindot()` using `bins = 30`. Pick better value with `binwidth`.
```




```
model1 <- glm(y ~ ap + week, data = bacteria, family = binomial())

summary(model1)
```

```
##
## Call:
## glm(formula = y ~ ap + week, family = binomial(), data = bacteria)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.2875   0.3895   0.5400   0.7011   1.0179
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  1.65020    0.31090   5.308 1.11e-07 ***
## ap           0.89034    0.37844   2.353  0.01864 *
## week        -0.11479    0.04395  -2.612  0.00901 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 217.38  on 219  degrees of freedom
## Residual deviance: 204.95  on 217  degrees of freedom
## AIC: 210.95
##
## Number of Fisher Scoring iterations: 4
```

```
compliance <- bacteria[ which(bacteria$hilo == 'hi'), ]
```

```
model2 <- glm(y ~ ap + week, data = compliance, family = binomial())

summary(model2)
```

```
##
## Call:
## glm(formula = y ~ ap + week, family = binomial(), data = compliance)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.4433   0.3335   0.4768   0.6460   0.9702
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  2.00466    0.47751   4.198 2.69e-05 ***
## ap           0.92832    0.54367   1.708   0.0877 .
## week        -0.13597    0.06415  -2.119   0.0341 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 105.538  on 121  degrees of freedom
## Residual deviance:  98.137  on 119  degrees of freedom
## AIC: 104.14
##
## Number of Fisher Scoring iterations: 5
```

```
basic = glm(y ~ ., data = bacteria, family = binomial())
summary(basic)
```

```
##
## Call:
## glm(formula = y ~ ., family = binomial(), data = bacteria)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -2.29134   0.00003   0.00004   0.47512   1.99622
##
## Coefficients: (4 not defined because of singularities)
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept)  2.550e+00  1.251e+00   2.038 0.041561 *
## app          1.924e+01  8.359e+03   0.002 0.998164
## hilolo      -1.562e+00  1.617e+00  -0.966 0.334074
## week        -2.127e-01  6.377e-02  -3.335 0.000852 ***
## IDX02       -2.525e-01  1.721e+00  -0.147 0.883337
## IDX03        2.081e+01  7.562e+03   0.003 0.997804
## IDX04        1.572e+00  1.127e+04   0.000 0.999889
## IDX05        1.572e+00  1.127e+04   0.000 0.999889
## IDX06        2.080e+01  8.359e+03   0.002 0.998015
## IDX07       -2.056e+00  1.541e+00  -1.335 0.181989
## IDX08       -1.924e+01  8.359e+03  -0.002 0.998164
## IDX09        1.572e+00  1.127e+04   0.000 0.999889
## IDX10        1.879e+01  1.221e+04   0.002 0.998772
## IDX11        1.059e-02  1.127e+04   0.000 0.999999
## IDX12        4.367e-01  1.470e+00   0.297 0.766359
## IDX13        1.059e-02  1.127e+04   0.000 0.999999
## IDX14       -1.909e+01  8.359e+03  -0.002 0.998178
## IDX15        2.081e+01  7.562e+03   0.003 0.997804
## IDX16        7.000e-16  1.677e+00   0.000 1.000000
## IDX17       -1.924e+01  8.359e+03  -0.002 0.998164
## IDX18       -4.944e-01  1.463e+00  -0.338 0.735345
## IDX19       -1.880e+01  8.359e+03  -0.002 0.998206
## IDX20        1.924e+01  8.359e+03   0.002 0.998164
## IDX21        1.059e-02  1.127e+04   0.000 0.999999
## IDY01        1.059e-02  1.127e+04   0.000 0.999999
## IDY02       -1.973e+01  8.359e+03  -0.002 0.998117
## IDY03       -1.125e+00  1.542e+00  -0.729 0.465772
## IDY04       -1.372e+00  1.633e+00  -0.841 0.400593
## IDY05       -2.176e+01  8.359e+03  -0.003 0.997923
## IDY06        1.647e+00  1.184e+04   0.000 0.999889
## IDY07        4.367e-01  1.470e+00   0.297 0.766359
## IDY08       -1.672e-01  1.709e+00  -0.098 0.922058
## IDY09        1.710e+00  1.189e+04   0.000 0.999885
## IDY10        1.562e+00  1.617e+00   0.966 0.334074
## IDY11        1.059e-02  1.127e+04   0.000 0.999999
## IDY12        1.850e+01  1.238e+04   0.001 0.998808
## IDY13       -1.505e+00  1.612e+00  -0.933 0.350643
## IDY14        9.411e-02  1.267e+04   0.000 0.999994
## IDZ01        2.089e+01  9.524e+03   0.002 0.998250
## IDZ02        1.059e-02  1.127e+04   0.000 0.999999
## IDZ03        1.925e+01  7.562e+03   0.003 0.997969
## IDZ05       -2.036e+01  8.359e+03  -0.002 0.998057
## IDZ06       -2.125e+00  1.890e+00  -1.124 0.260955
```

```
## IDZ07        4.367e-01  1.470e+00   0.297 0.766359
## IDZ09        1.932e+01  8.389e+03   0.002 0.998162
## IDZ10        1.059e-02  1.127e+04   0.000 0.999999
## IDZ11        2.081e+01  7.562e+03   0.003 0.997804
## IDZ14        5.062e-16  1.677e+00   0.000 1.000000
## IDZ15        2.038e+01  1.006e+04   0.002 0.998383
## IDZ19        1.925e+01  7.562e+03   0.003 0.997969
## IDZ20               NA         NA      NA       NA
## IDZ24       -1.125e+00  1.542e+00  -0.729 0.465772
## IDZ26               NA         NA      NA       NA
## trtdrug             NA         NA      NA       NA
## trtdrug+            NA         NA      NA       NA
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 217.38  on 219  degrees of freedom
## Residual deviance: 118.51  on 169  degrees of freedom
## AIC: 220.51
##
## Number of Fisher Scoring iterations: 19
```