

# 3\_variations\_on\_flow\_n\_circle\_diagrams

Diardano Raihan

2022-07-21

## Import the Libraries and Dataset

```
library(tidyverse)

## -- Attaching packages ----- tidyverse 1.3.1 --

## v ggplot2 3.3.6      v purrr  0.3.4
## v tibble  3.1.7      v dplyr  1.0.9
## v tidyr   1.2.0      v stringr 1.4.0
## v readr   2.1.2      v forcats 0.5.1

## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()

library(knitr)
library(ggthemes)
library(RColorBrewer)
```

## Flows

### Alluvial Diagrams

- A good visualization to show parts of a whole world connections between different categories and data.
- Let's create some fake data about student performance in classes.
  - 21 students are divided in 3 groups, they are men and women, and they get High Pass, Pass, or Fail.
- Reference:

```
group1 <- tibble(groupid = groupid <- rep("group1", 7),
                 studentID = sample(seq(from = 1, to = 20), 7),
                 gender = sample(c("M", "F"), 7, replace = TRUE),
                 grades = sample(c("High Pass", "Pass", "Fail"), 7, replace = TRUE))

group2 <- tibble(groupid = groupid <- rep("group2", 7),
```

```

      studentID = sample(seq(from = 21, to = 30), 7),
      gender = sample(c("M", "F"), 7, replace = TRUE),
      grades = sample(c("High Pass", "Pass", "Fail"), 7, replace = TRUE))

group3 <- tibble(groupid = rep("group3", 7),
  studentID = sample(seq(from = 31, to = 40), 7),
  gender = sample(c("M", "F"), 7, replace = TRUE),
  grades = sample(c("High Pass", "Pass", "Fail"), 7, replace = TRUE))

students <- bind_rows(group1, group2, group3)
students

```

```

## # A tibble: 21 x 4
##   groupid studentID gender grades
##   <chr>      <int> <chr>  <chr>
## 1 group1      10 M      Pass
## 2 group1      14 F      High Pass
## 3 group1      16 F      Pass
## 4 group1       3 F      Fail
## 5 group1      12 M      Fail
## 6 group1      20 F      Fail
## 7 group1      18 F      Fail
## 8 group2      21 M      High Pass
## 9 group2      28 M      High Pass
## 10 group2     22 M      High Pass
## # ... with 11 more rows

```

```

# summarize data
students_table <- students %>%
  group_by(groupid, gender, grades) %>%
  count()

```

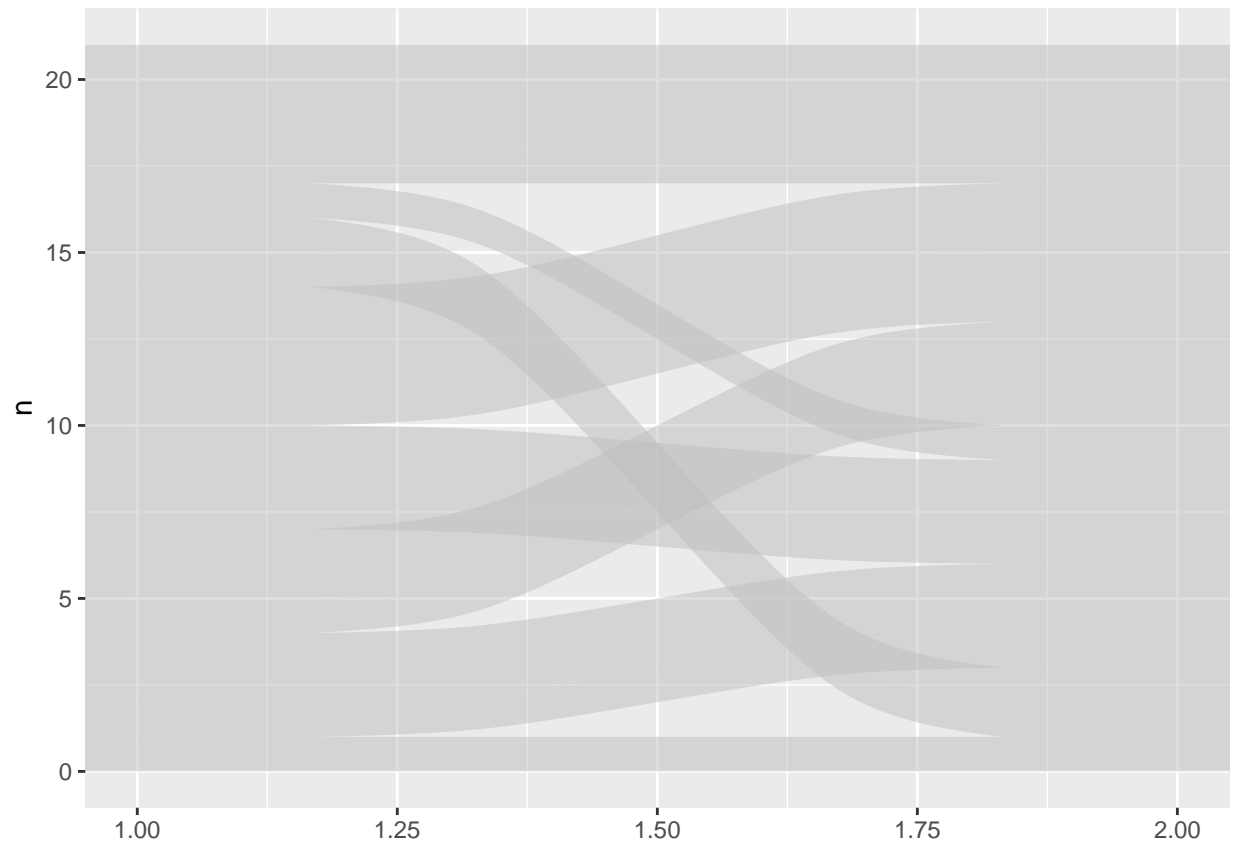
- Now, let's draw the alluvial framework

```

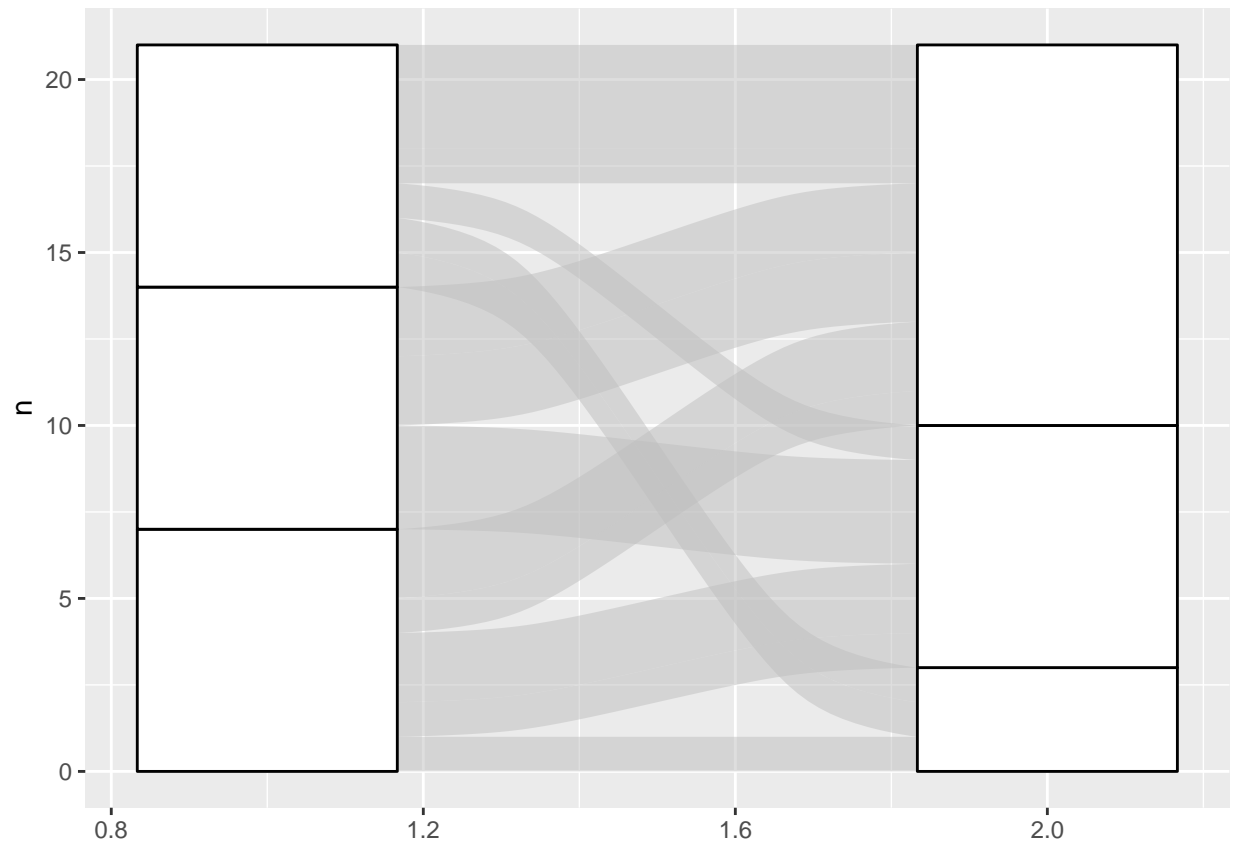
# install.packages("ggalluvial")
library(ggalluvial)

# barebones alluvial
ggplot(students_table,
  aes(axis1 = groupid,
      axis2 = grades,
      y = n))+
  geom_alluvium()

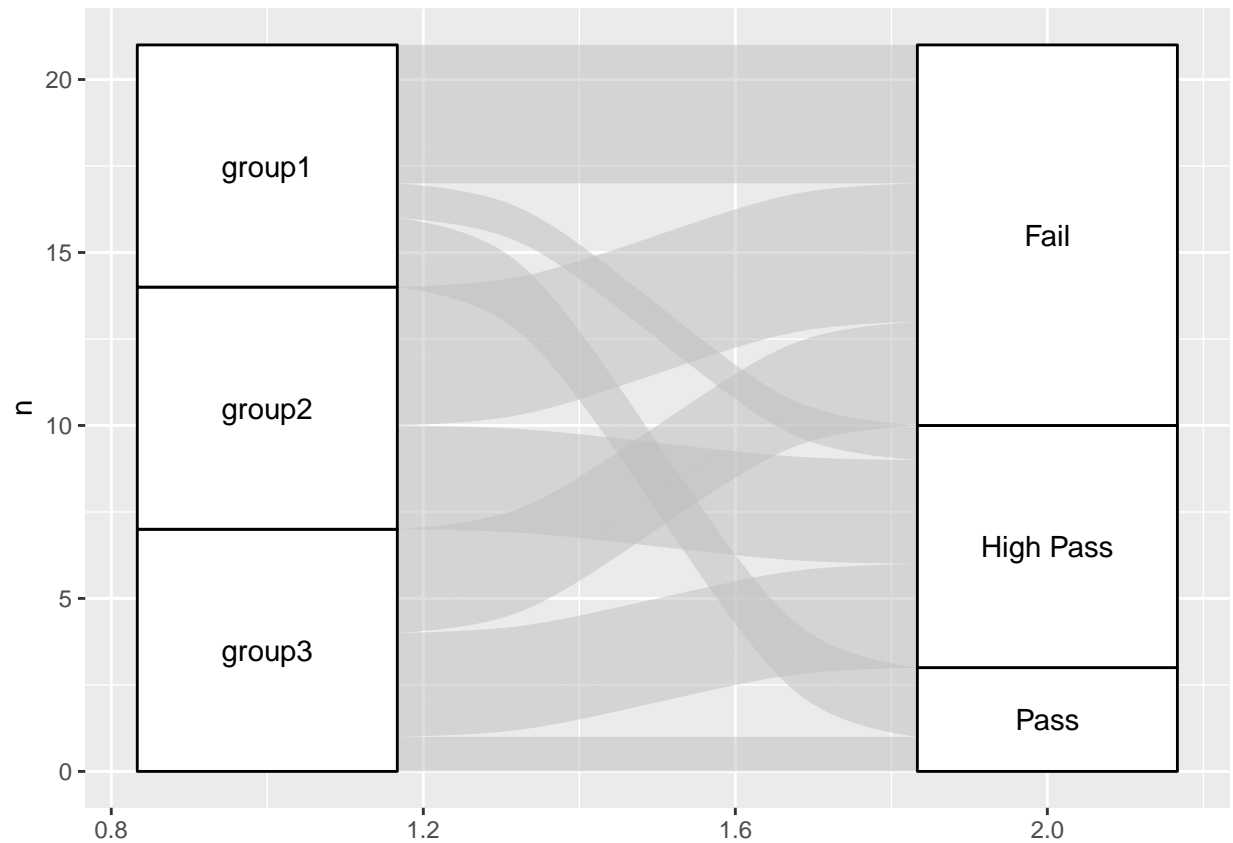
```



```
# add strata
ggplot(students_table,
  aes(axis1 = groupid,
      axis2 = grades,
      y = n)) +
  geom_alluvium() +
  geom_stratum()
```



```
# add labels
ggplot(students_table,
  aes(axis1 = groupid,
      axis2 = grades,
      y = n)) +
  geom_alluvium() +
  geom_stratum() +
  geom_text(stat = "stratum",
    aes(label = after_stat(stratum)))
```



```
# change colors
ggplot(students_table,
  aes(axis1 = groupid,
      axis2 = grades,
      y = n))+
  geom_alluvium(aes(fill = gender))+
  geom_stratum()+
  geom_text(stat = "stratum",
    aes(label = after_stat(stratum)))+
  scale_fill_manual(values = c("purple", "green"))
```

