ggalluvial Cheat Sheet

by seleven via cheatography.com/139867/cs/29584/

GGALLUVIAL ---- Cheatsheet

Introduction to ggalluvial

The **ggalluvial** package is a ggplot2 extension for producing alluvial plots.
Alluvial plots use variable-width ribbons and stacked bar plots to represent multi-dimensional or repeated-measures data with categorical or ordinal variables.
There are two types of alluvial format: **Alluvial (Wide) Format** & **Lodes (Long) Format**

More information

https://cran.r-project.org/web/packages/ggalluvial/ggalluvial.pdf

Five Essential Components

1. AIXS	A dimension (variable) along which the data are vertically grouped at a fixed horizontal position.
2. ALLUVIUM	Horizontal (x-) splines called alluvia span the width of the plot.
3. STRATUM	The groups at each axis are depicted as opaque blocks called strata.
4. LODE	The alluvia intersect the strata at lodes.
5. FLOW	The segments of the alluvia between pairs of adjacent axes are flows.

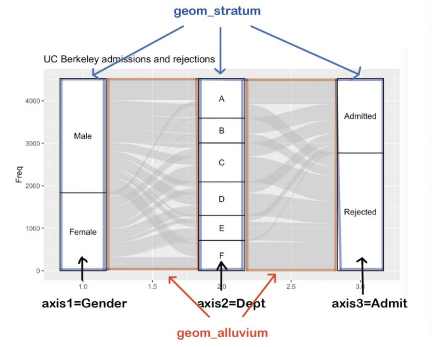
Basic Alluvial Wide Format

Discription	Example
load packages	library(ggalluvial)
basic ggplot	gg <- ggplot(as.data.frame (UCBAdmissions), aes(y = Freq, axis1 = Gender, axis2 = Dept, axis3 = Admit))
add alluvium	+ geom_alluvium()
add stratum	+ geom_stratum()
add text	+ geom_text(stat = "stratum", aes(label = paste(after_stat(stratum))))
add title	+ ggtitle("UC Berkeley admissions and rejections")

The dataset **"UCBAdmissions"** is an aggregate data on applicants to graduate school at Berkeley for the six largest departments in 1973 classified by admission and sex.
It is a **3-dimensional array** resulting from cross-tabulating 4526 observations on 3 variables.

No	Name	Levels
1	Admit	Admitted, Rejected
2	Gender	Male, Female
3	Dept	A, B, C, D, E, F

Graph of Wide Format

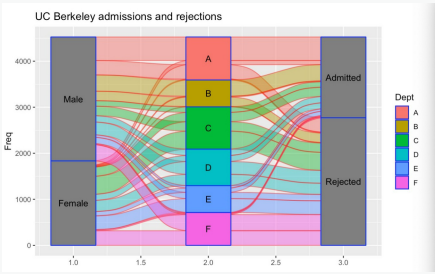


Change Color

the border of alluvium	geom_alluvium(color="red")
the border of stratum	geom_stratum(color="blue")
the fill of alluvium	geom_alluvium(aes(fill=Dept)) geom_alluvium(aes(fill=Gender)) geom_stratum(aes(fill=Admit))
the fill of stratum	geom_stratum(aes(fill=Dept)) geom_stratum(aes(fill=Gender)) geom_stratum(aes(fill=Admit))

Setting the different fills of alluvium and stratum can help analysts easily analyze the data from different aspects.

Graph after changing Color & Fill by Dept



If using **"fill=Dept"**, it means we are using colors grouped by each department.
It can help analyst to see the **formation** of each department: how many males and females in each department.
Also it shows how many people in each department are admitted and rejected.

By **seleven**
cheatography.com/seleven/

Published 29th October, 2021.
Last updated 29th October, 2021.

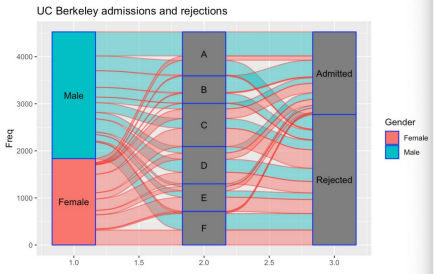
Sponsored by **ApolloPad.com**
Everyone has a novel in them. Finish Yours!
https://apollopad.com

Cheatography

ggalluvial Cheat Sheet

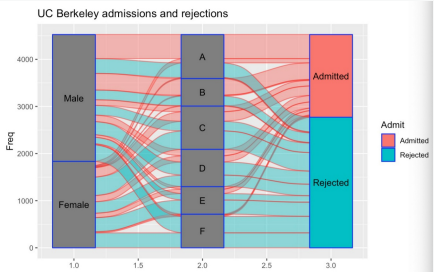
by seleven via cheatography.com/139867/cs/29584/

Graph after changing Color & Fill by Gender



If using **"fill=Gender"**, it means we are using colors grouped by different gender.
It can help analyst to see how many males and females apply for each department and finally admitted or rejected.

Graph after changing Color & Fill by Admit

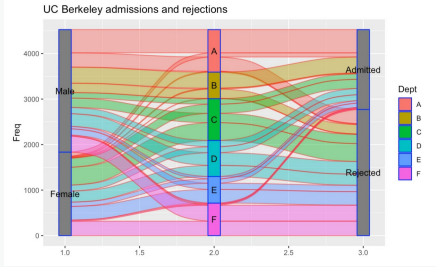


If using **"fill=Admit"**, it means we are using colors grouped by admitted or rejected.
It can help analyst to see the **formation** admitted students: how many admitted students are from each department and of different gender.

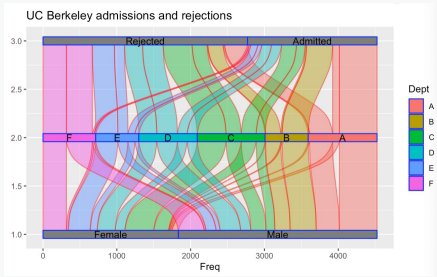
Change Width

the width of alluvium	geom_alluvium(color="red", aes(fill=Dept),width=1/12)
the width of stratum	geom_stratum(color="blue", aes(fill=Dept),width=1/12)

Graph after changing Width

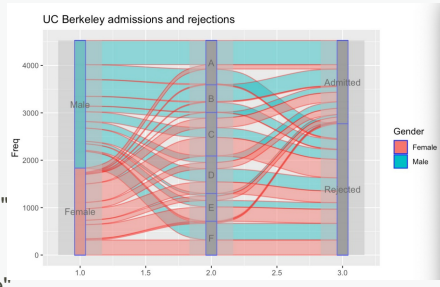


Flip Coordinates



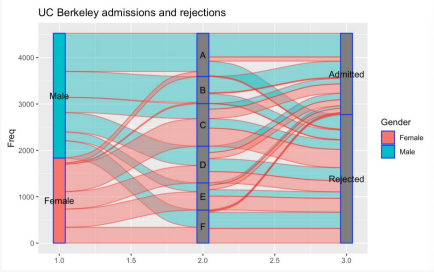
Adding **coord_flip()**
Flip cartesian coordinates so that horizontal becomes vertical, and vertical, horizontal. This is primarily useful for converting geoms and statistics which display y conditional on x, to x conditional on y.

Adding lode



Adding **geom_lode()**

geom_alluvium vs geom_flow



The graph is using **geom_flow**.
We can see the difference between **geom_alluvium** and **geom_flow**.
After we use "flow", all males apply for department A came together, which is also the same as other departments. It makes the graph much clearer than before since there is less cross alluviums between each axises.

More coding help

Description	Example
Adding the names of each axis	+scale_x_discrete(limits = c("Gender", "Dept","Admit"))
Changing the fill of stratum	+scale_fill_brewer(type = "qual", palette = "Set1")

Basic Lodes (Long) Format

Description	Example
Convert data to Lodes format	to_lodes_form(as.data.frame(UCBAdmissions),axes = 1:3,id = "Cohort")
load data	data(majors) majors$curriculum <- as.factor(majors$curriculum)

By **seleven**
cheatography.com/seleven/

Published 29th October, 2021.
Last updated 29th October, 2021.

Sponsored by **ApolloPad.com**
Everyone has a novel in them. Finish Yours!
https://apollopad.com

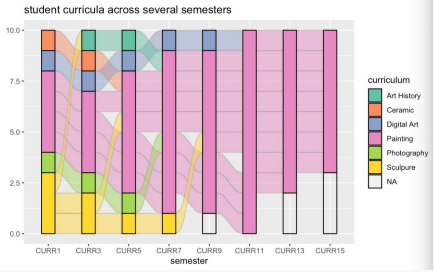
Basic Lodes (Long) Format (cont)

basic ggplot	ggplot(majors,aes(x = semester, stratum = curriculum, alluvium = student,fill = curriculum, label = curriculum))
add flow	+geom_flow(stat = "alluvium", lode.guidance = "frontback",color = "darkgray") +
add stratum	+geom_stratum()
add title	+ggtitle("student curricula across several semesters")

The long format requires an additional indexing column that links the rows corresponding to a common cohort.

The data follows the major curricula of 10 students across 8 academic semesters.
Missing values indicate undeclared majors.
A data frame with 80 rows and 3 variables:
1. student: student identifier
2. semester: character tag for odd-numbered semesters
3. curriculum: declared major program

Graph of Lodes Format



This graph clearly shows a set of students' academic curricula over the course of several semesters.
The lode format gives us the option to aggregate the flows between adjacent axes, which may be appropriate when the transitions between adjacent axes are of primary importance.

Sponsored by **ApolloPad.com**
Everyone has a novel in them. Finish Yours!
https://apollopad.com