

Extensions to ggplot2 (ggpubr, gggrides, ...)

Sang Yeol Lee

DataBreak
ShinyKorea-Meetup

September 26, 2019

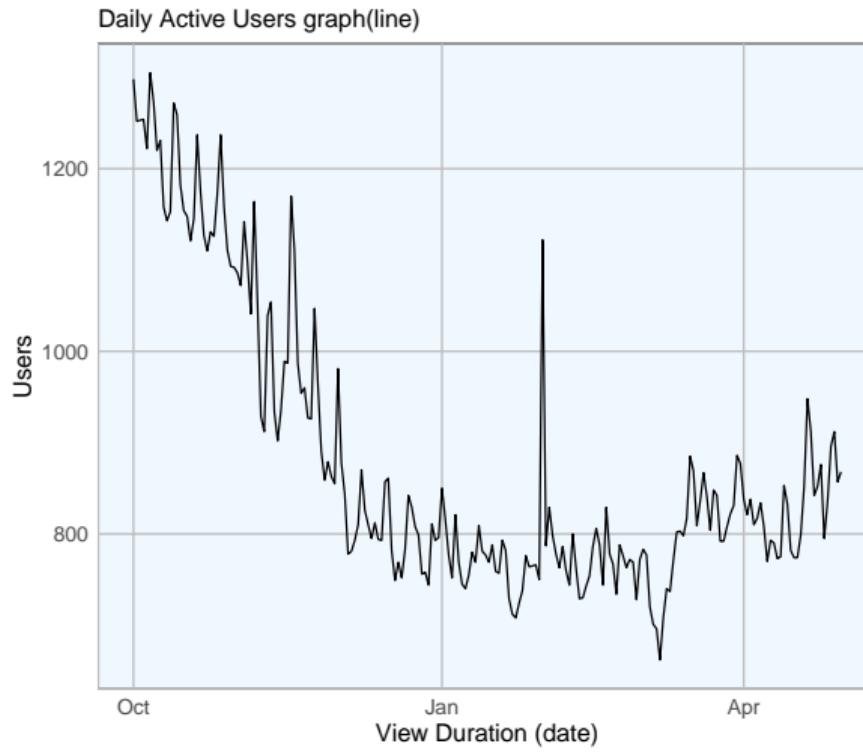
1. Intro

- 데이터뽀개기 운영 을 하고 있습니다. 커뮤니티 운영하느라 공부는 소홀해지고 있네요…
- 데이터 탐색 시각화는 R로 하고 머신러닝 및 서비스 적용하기 위한 개발은 python으로 합니다.
- 요즘 관심사는 추천시스템이고 pyspark-Mllib 통해 모델링하고 개발하고 있습니다.
- 오늘은 12 Extensions to ggplot2 for More Powerful R Visualization 내용을 소개드릴 예정입니다.
그 중에서 최근 제가 사용했던 ggpibr, ggridges에 대해서 소개드릴게요. 원본
- ggplot2을 다루는 분들은 환영입니다.!

2. ggplot2

- ggplot2는 2007년에 소개된 이래 세계에서 가장 많이 다운로드되고 널리 사용되는 R 패키지 중 하나입니다.
- 패키지는 Leland Wilkinson의 주요 텍스트인 The Grammar of Graphics에서 영감을 얻어 스케일과 레이어와 같은 시맨틱 컴포넌트로 그래프를 나누고 프리젠테이션과 데이터를 분리합니다.
- R 커뮤니티는 지난 10년 동안 ggplot2 용 확장 패키지를 구축하는 작업에 열심히 노력했습니다.

2. ggplot2 - code



2. ggplot2 - code (2)

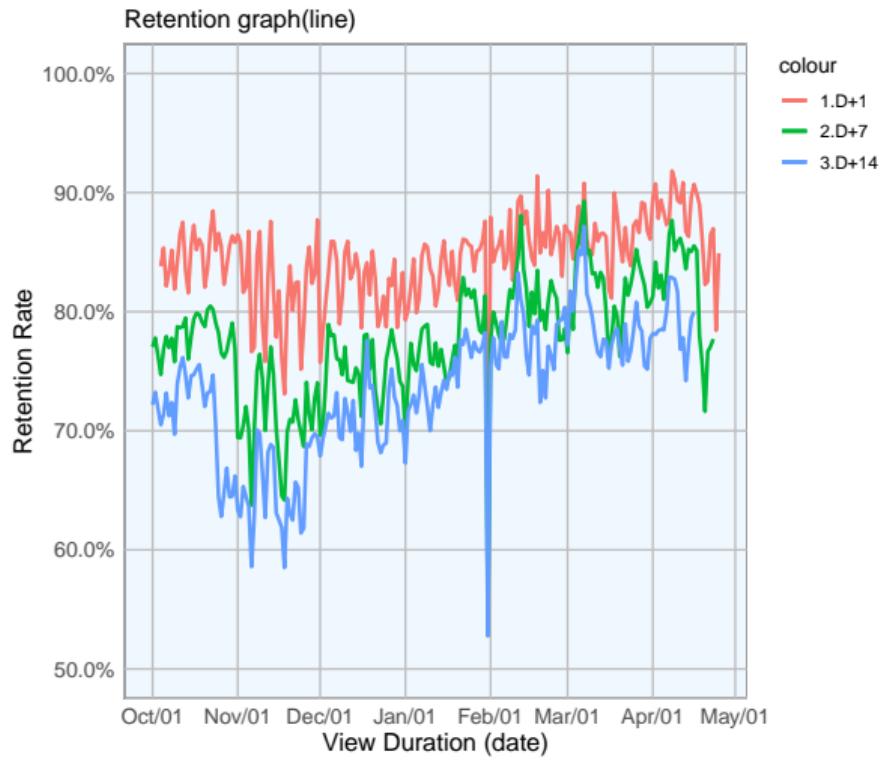
	first_date	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	2018-10-01	1298	1083		1042	1021	1017	1005	1000	981	948	929	929	955	931	937
2	2018-10-02	1252		1044	1019	1010	999	988	974	941	939	937	952	936	927	917
3	2018-10-04	1254	1051	1040	1024	1000	991	955	937	941	963	940	932	926	893	884
4	2018-10-05	1222	1043	1008	994	987	961	933	939	960	939	917	913	883	885	872
5	2018-10-06	1305	1072	1029	1015	967	947	951	1017	972	941	934	906	886	876	955
6	2018-10-07	1273	1060	1034	980	950	943	987	980	938	925	902	893	877	929	907
7	2018-10-08	1220	1039	987	957	954	978	957	949	927	903	894	893	928	888	883
8	2018-10-09	1231	1008	970	961	979	972	945	933	907	898	877	925	902	877	858
9	2018-10-10	1158	974	948	961	949	934	927	912	889	889	909	883	873	847	856
10	2018-10-11	1143	989	977	966	954	932	921	899	887	912	893	879	860	862	861

Showing 1 to 10 of 211 entries

Previous

1 2 3 4 5 ... 22 Next

2. ggplot2 - code (3)

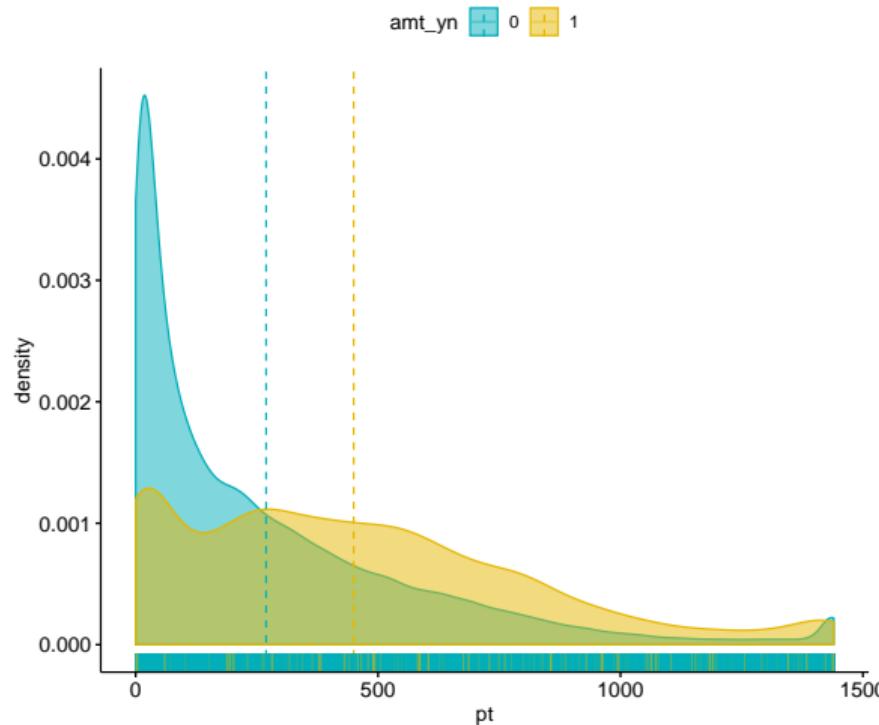


3. ggpubr

- ggplot2 기반 그래프 생성을 쉽게 할 수 있는 라이브러리 (논문에 출판 가능한 그림을 이쁘게 그릴 수 있음!!!)
- 쉽게 주석을 달거나, 정렬하거나, 색상 및 레이블과 같은 세부적인 매개 변수 컨트롤이 쉬움
- Automatically add p-values and significance levels to box plots, bar plots, line plots, and more
- 참고 레퍼런스
 - Facilitating Exploratory Data Visualization: Application to TCGA Genomic Data
 - Add P-values and Significance Levels to ggplots
 - Perfect Scatter Plots with Correlation and Marginal Histograms
 - Plot Means/Medians and Error Bars
 - Bar Plots and Modern Alternatives
 - Add Text Labels to Histogram and Density Plots
 - ggplot2 - Easy Way to Mix Multiple Graphs on The Same Page
 - ggplot2 - Easy Way to Change Graphical Parameters
 - Create and Customize Multi-panel ggplots: Easy Guide to Facet

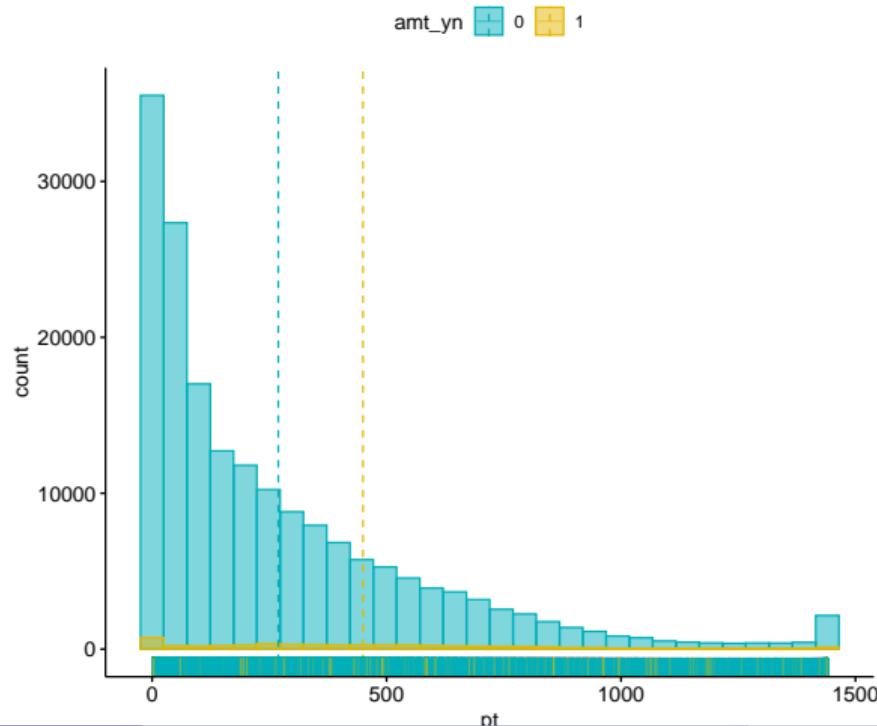
3. ggpubr (1)

- Distribution



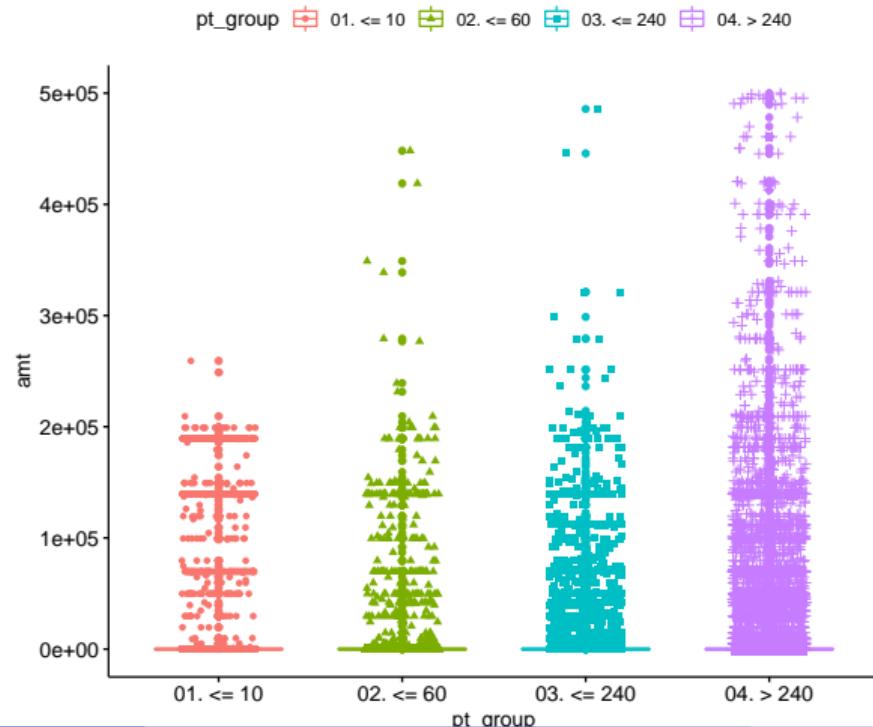
3. ggpubr (2)

- `gghistogram`



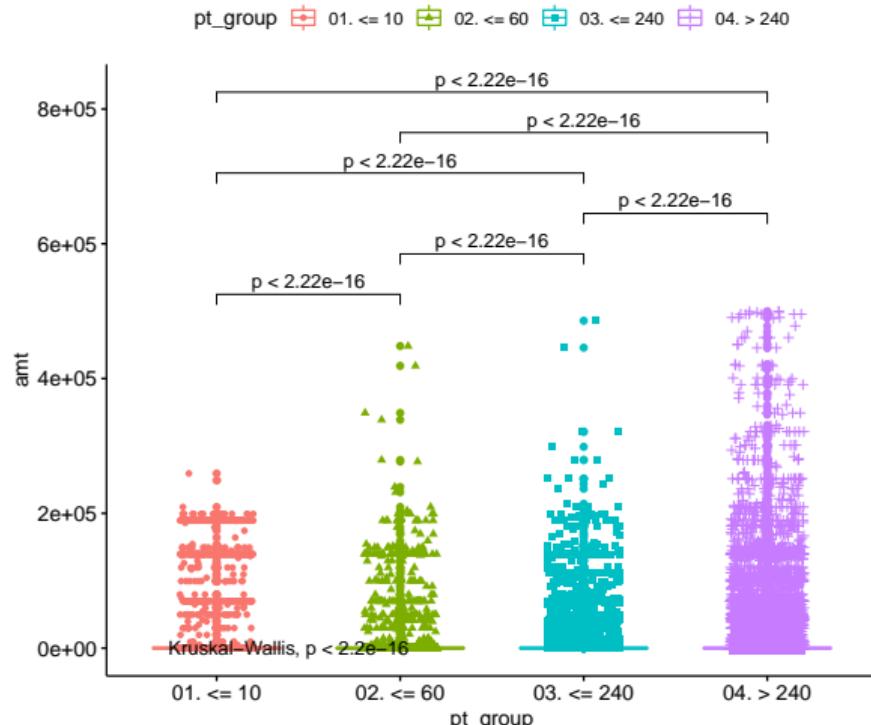
3. ggpubr (3)

- Box plots and violin plots



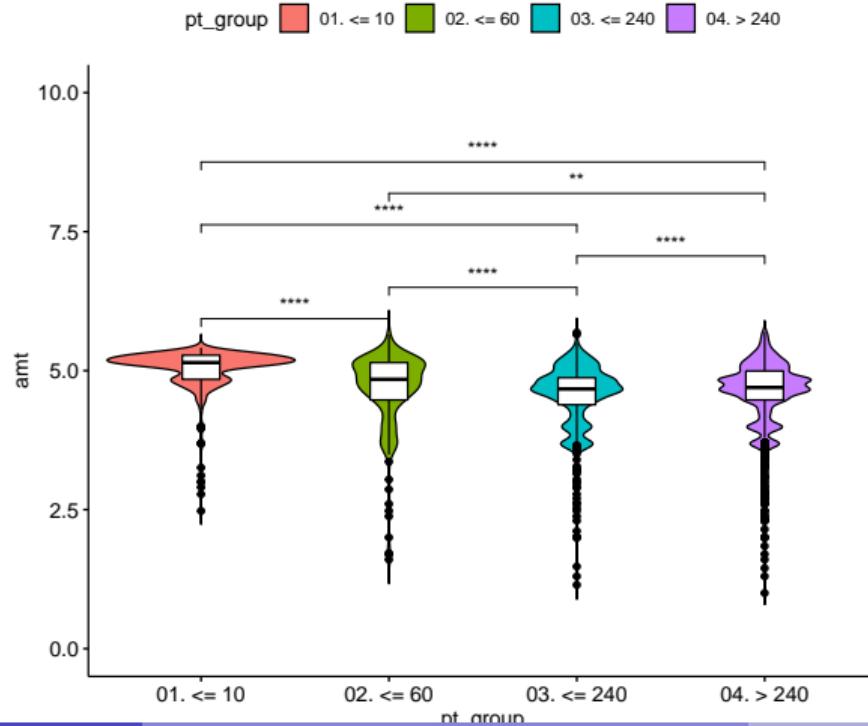
3. ggpubr (4)

- Box plots and violin plots



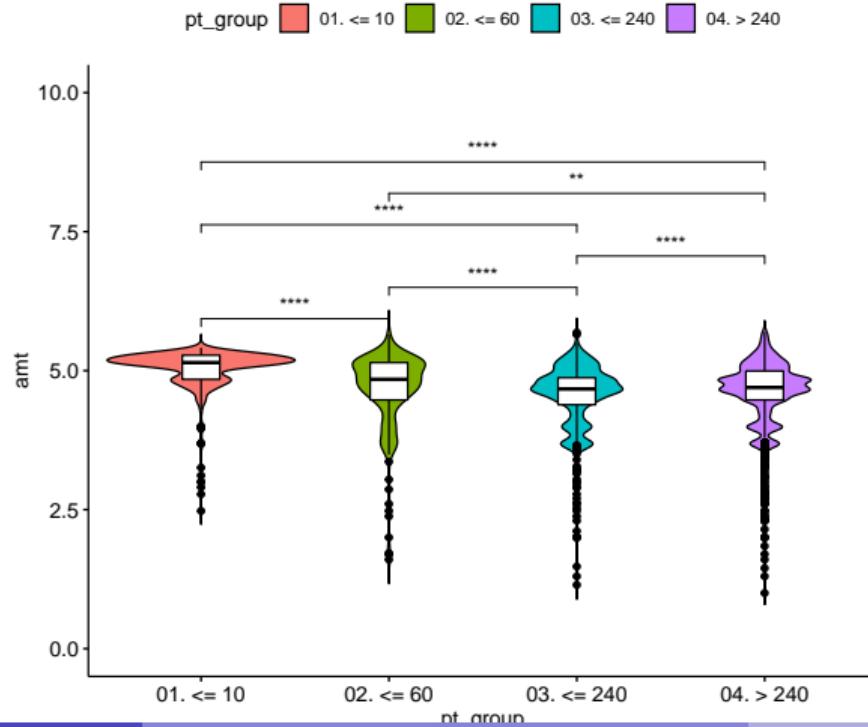
3. ggpubr (5)

- Box plots and violin plots



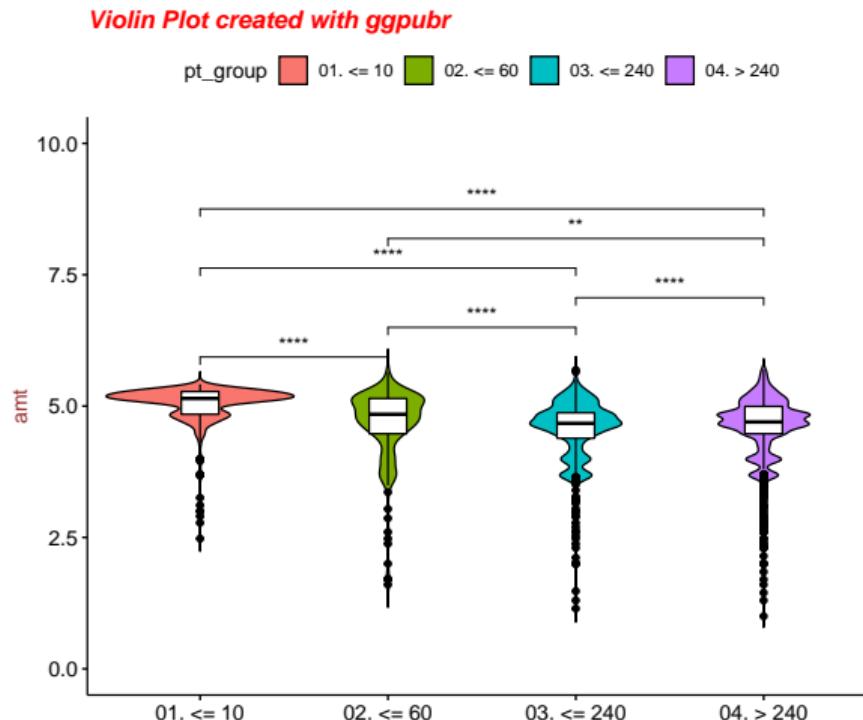
3. ggpubr (6)

- Box plots and violin plots



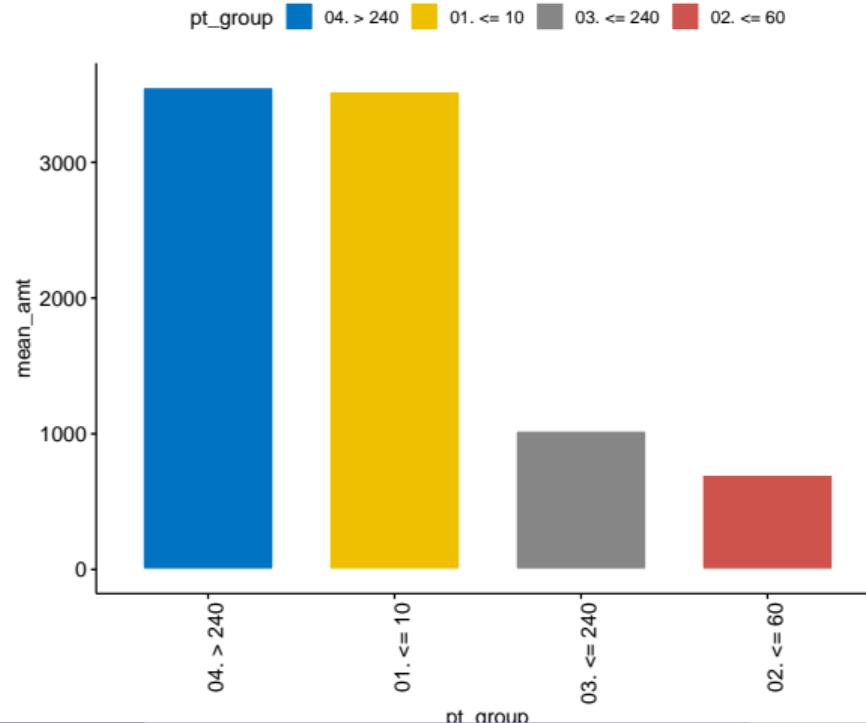
3. ggpubr (7)

- Change titles and axis labels



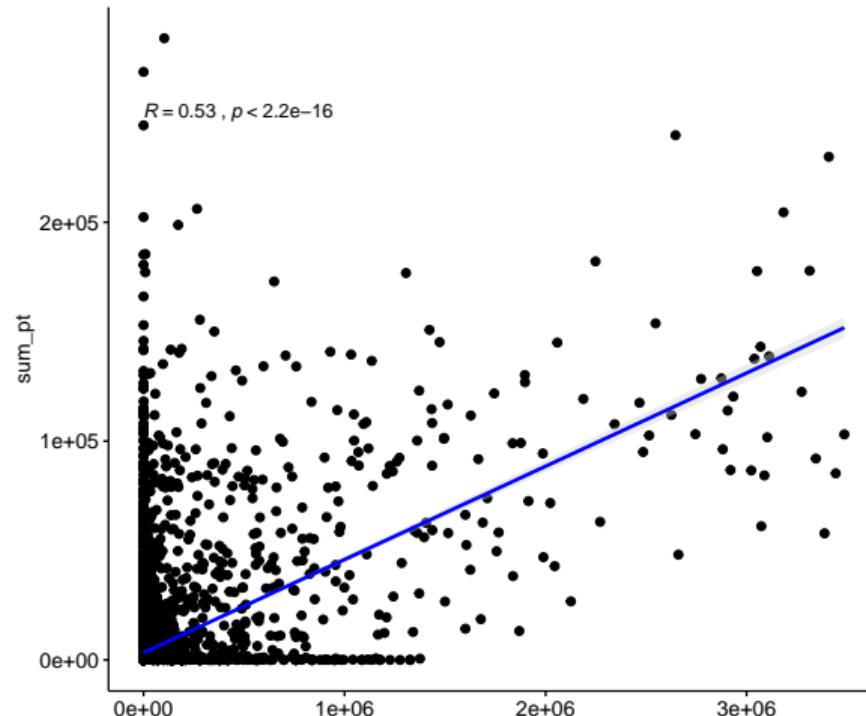
3. ggpubr (8)

- Ordered bar plots



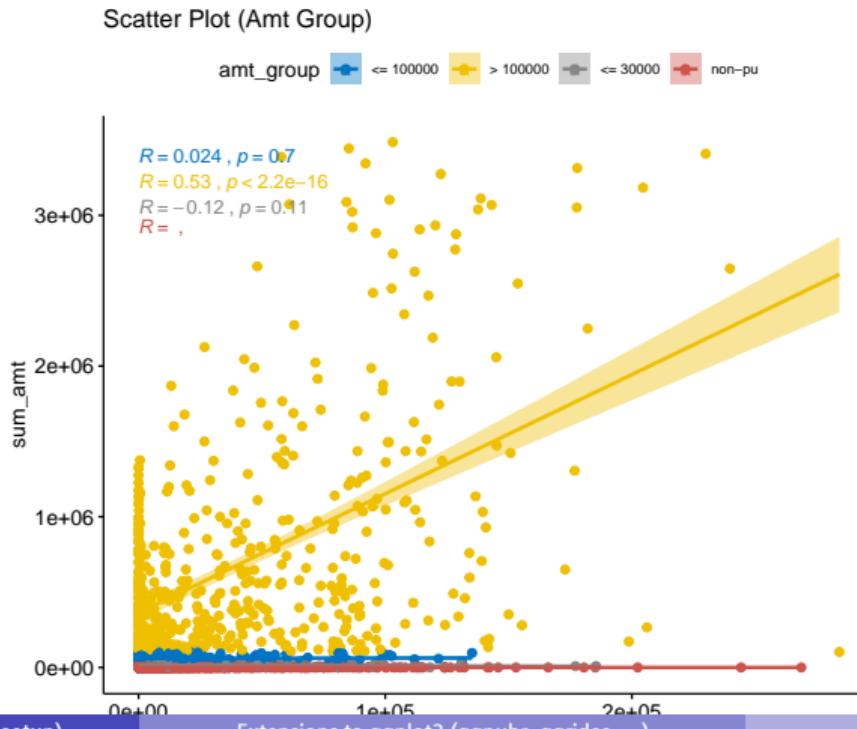
3. ggpubr (9)

- Perfect Scatter Plots with Correlation and Marginal Histograms



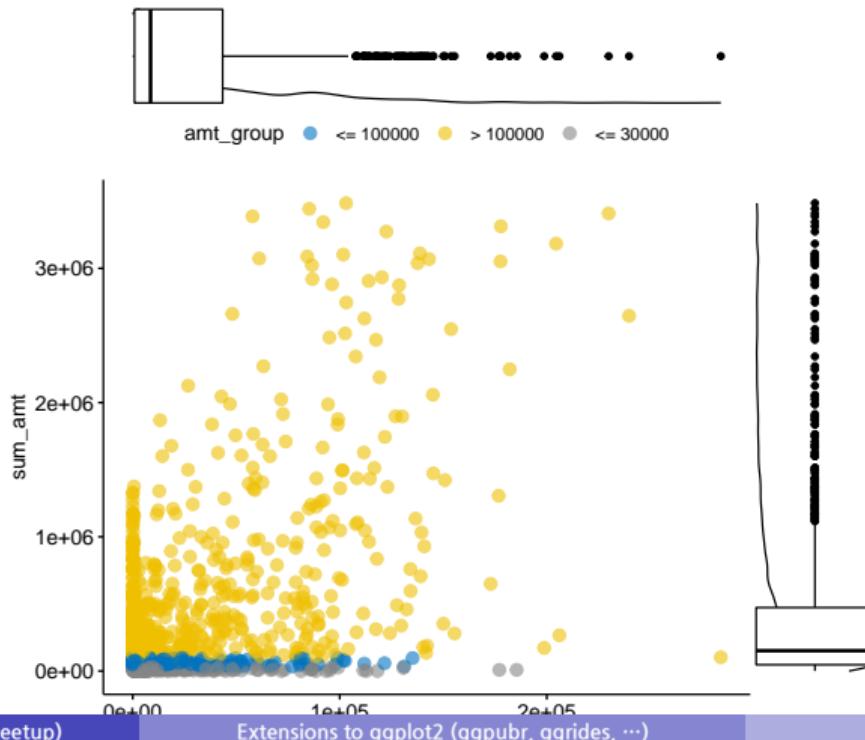
3. ggpubr (10)

- Color by groups (Grouping variable: 결제금액_group). To add a correlation coefficient per group, specify the grouping variable using the mapping function aes() as follow.



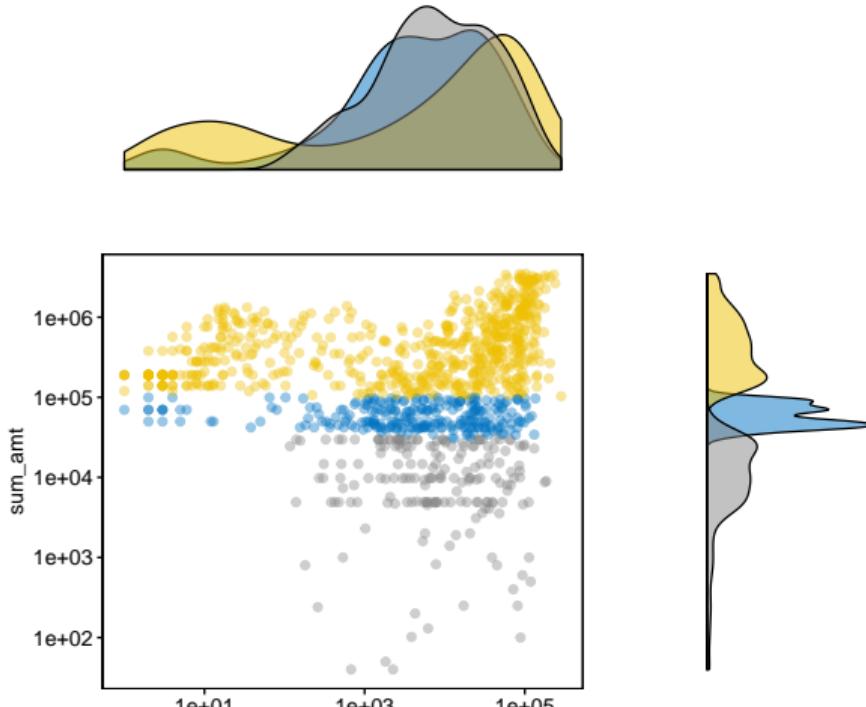
3. ggpubr (11)

- marginal 플롯 추가 (ggExtra 패키지의 ggMarginal() 함수는 분산 히스토그램, 밀도 또는 상자 그림을 산포도에 쉽게 추가하는 데 사용할 수 있습니다.) (ggExtra 패키지 설치 필요)



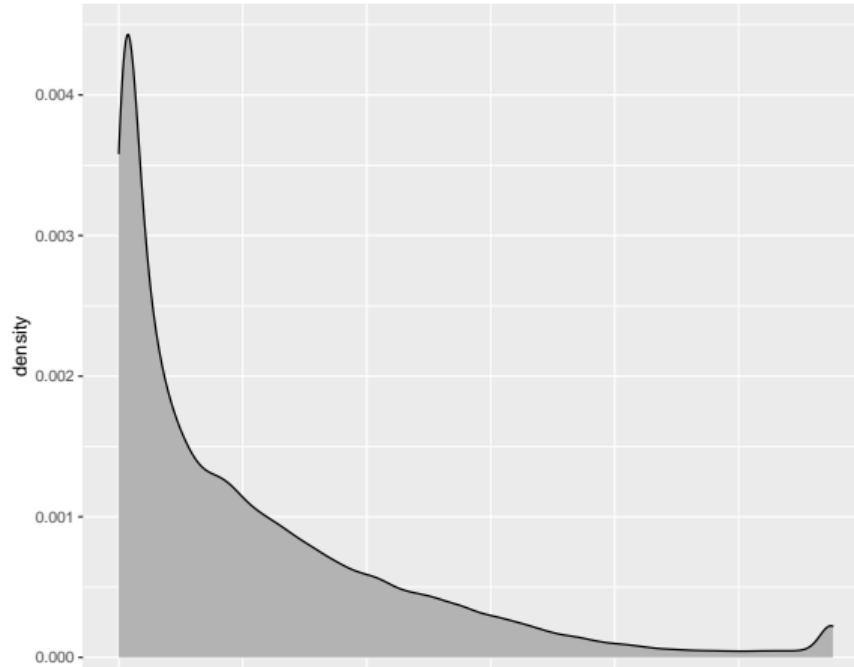
3. ggpubr (12)

-ggExtra의 한 가지 제한은 산점도와 marginal plots에서 여러 그룹을 처리 할 수 없다는 것. 아래의 R 코드에서 cowplot 패키지를 사용하는 솔루션을 활용



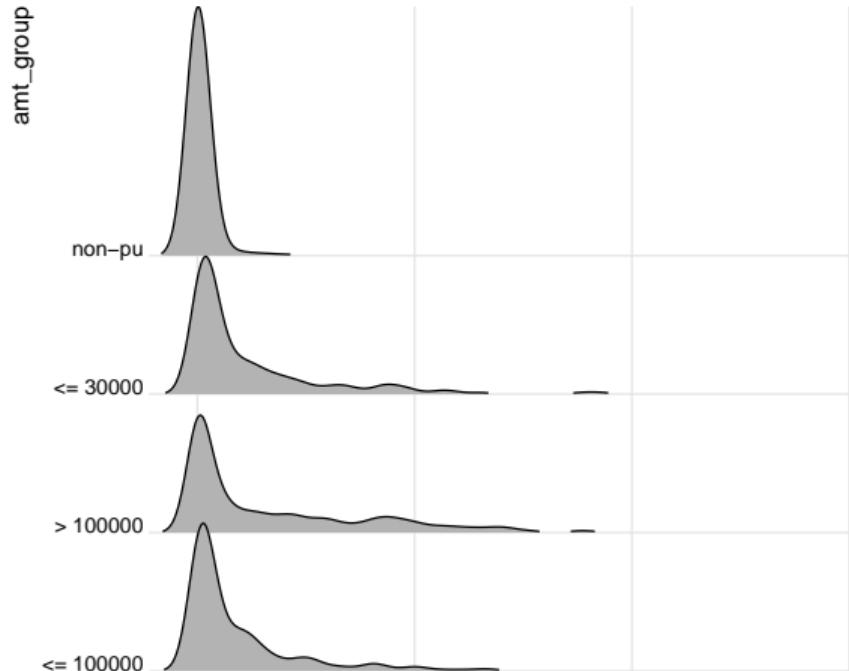
4. gggrides (1)

- Ridgeline(능선) 도표는 시간 또는 공간에 따른 분포의 변화를 시각화 하는 편리한 방법을 제공합니다.
- reference



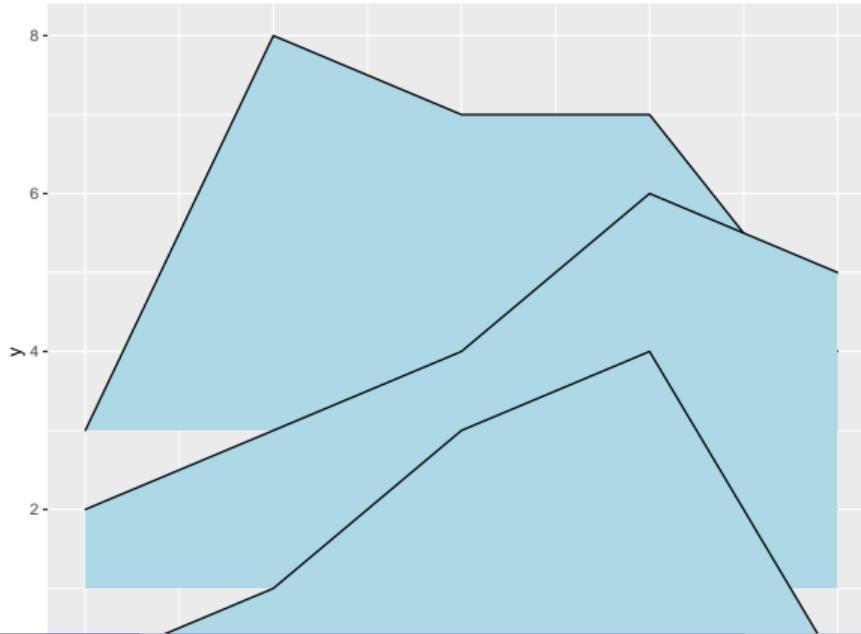
4. gggrides (2)

-geom_density_ridges는 유명한 Joy Division 앨범 Unknown Pleasures의 표지와 같이 여러 밀도 플롯을 엇갈리게 배열합니다. Unknown Pleasures



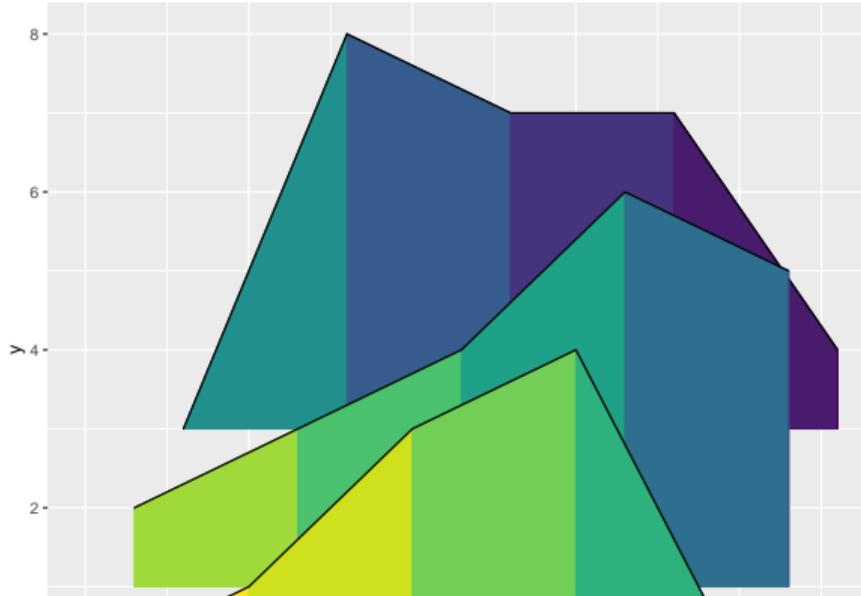
4. gggrides (2)

- Ridgeline 도표는 시간 또는 공간에 따른 분포의 변화를 시각화하는 편리한 방법을 제공합니다. 이 패키지는 'ggplot2'에서 이러한 플롯을 생성할 수 있습니다.
- y와 높이에 매핑 된 데이터는 동일한 단위여야 함. 높이를 상대적으로 조정하려면 stat = "identity"로 geom_density_ridges를 사용할 수 있습니다



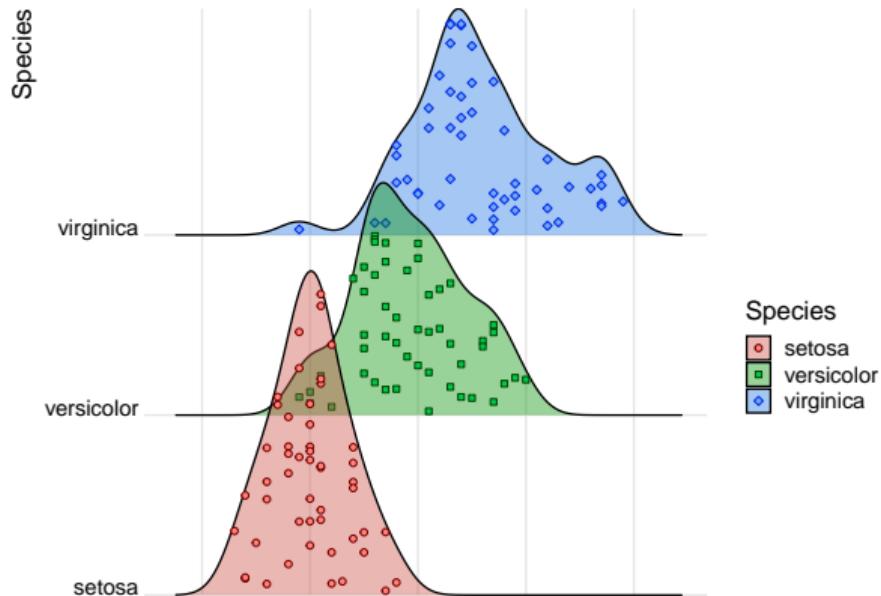
4. gggrides (3)

- geom_ridgeline_gradient 및 geom_density_ridges_gradient ggms_ridgeline_gradient 및 geom_density_ridges_gradient는 채우기 미학이 x 축을 따라 변할 수 있다는 점을 제외하고 geom_ridgeline 및 geom_density_ridges와 동일하게 작동합니다.
- 꼭 필요한 경우가 아니라면 사용하지 마십시오. R 그래픽 시스템의 한계로 인해 그라디언트 채우기에 투명도 (α)를 비활성화 해야함



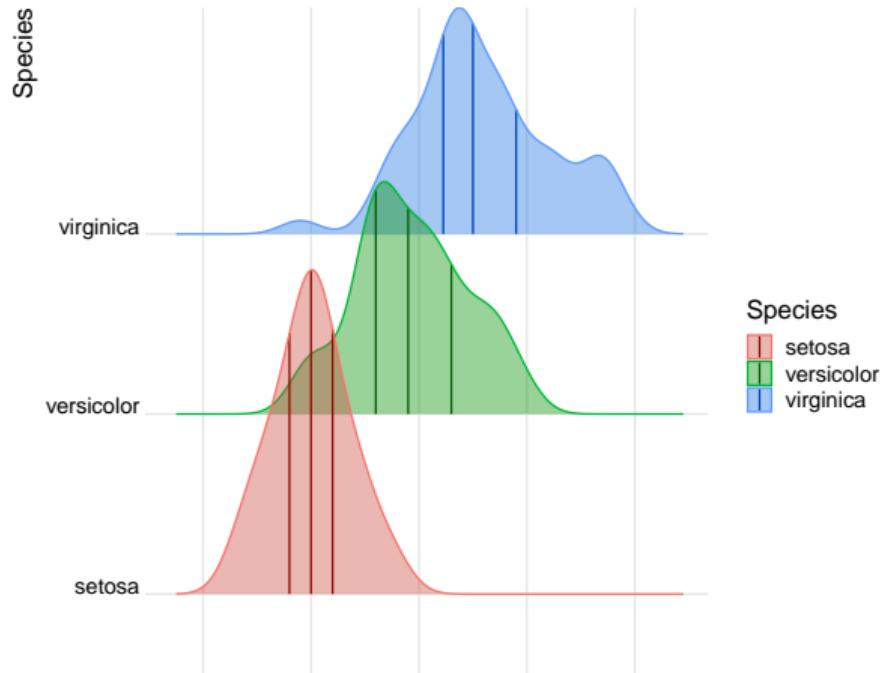
4. gggrides (4)

- point_color, point_fill, point_size와 같이 포인트 미학에 적용할 수 있는 다양한 스케일입니다. 개별 스케일은 모두 기존 표준 ggplot2 스케일과 사용법이 동일하며 이름만 다릅니다.



4. gggrides (5)

- point 미학에 대한 특정 배율에 대해서는 `scale_point_color_hue()` 및 일반 이산 스케일에 대해서는 `scale_discrete_manual()`을 참조하십시오.



4. gggrides (6)

- 출력이 카운트 set가 아닌 히스토그램을 설명하는 능선이라는 점을 제외하고 stat_bin 함수와 동일하게 작동합니다.

