

大数据高端人才专项计划



東北大學
Northeastern University



HORIZON
昊宸科技





数据介绍:

这是一份电商用户的行为数据, 其中item_id为商品id, cate_id为仓库的id。

为这个数据添加字段名

```
colnames(df) <-
```

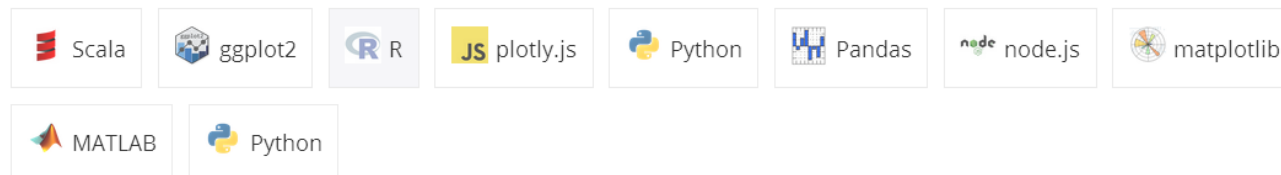
```
c("date", "item_id", "cate_id", "cate_level_id", "brand_id", "supplier_id", "pv_ipv", "cart_uv", "collect_uv", "cart_ipv")
```

20150628	300	36	4	657	294	33	19	1	1
20150130	300	36	4	657	294	391	119	16	14
20141021	300	36	4	657	294	53	33	5	4
20150213	300	36	4	657	294	110	54	2	2
20150212	300	36	4	657	294	134	55	3	2
20151106	300	36	4	657	294	30	10	7	4
20141031	300	36	4	657	294	77	44	5	5
20150211	300	36	4	657	294	126	72	4	4
20150210	300	36	4	657	294	119	62	2	2
20151107	300	36	4	657	294	12	7	2	2
20150722	300	36	4	657	294	98	31	5	4
20150629	300	36	4	657	294	28	18	0	0

- plotly是R语言中的一个包，基于开源的JS图形库plotly.js，用于创建交互式的基于网页的图形。
- plotly支持离线或在线使用。
- 支持多种图表类型。
- 多种方式分享分析结果，可以嵌入网页，生成pdf、png、svg等格式，创建自定义演示或仪表板。

Visualize Data,
Together


Plotly is the modern platform for agile business intelligence and data science.



图表类型



Basic


 Scatter plot

 Line plot


 Bar chart


 Area Chart


 Pie chart


 Animation **PRO**

Statistical


 Histogram

 Box plot

 2D Histogram **PRO**
Heatmap


 2D Histogram **PRO**
Contour Plot

Scientific


 Error bars


 Heatmap

 Contour

 Ternary plot **PRO**

Maps

 Choropleth **PRO**

 Atlas Maps **PRO**

 Satellite Maps **PRO**


3d Charts

 3D Scatter


 3D Surface

 3D Mesh

Financial

 Time Series

 Candlestick Chart

 OHLC Chart

安装plotly



- plotly 已经收录于cran 上

```
install.packages("plotly")
```

- 或者通过devtools从github上下载最新的版本

```
devtools::install_github("ropensci/plotly")
```

- 离线绘图

```
library(plotly)
p <- plot_ly(midwest, x = ~percollege, color = ~state, type = "box")
p
```

plotly的两种方式



`ggplotly()/plot_ly()`

`plot_ly()`函数将数据转换成plotly对象, `ggplotly()`函数将ggplot对象转换成plotly对象

plotly依赖于ggplot2

`plot_ly()`也是基于《grammar of graphics》

`plot_ly()`更加灵活、简洁



plotly图表主要通过`plotly::plot_ly`, `plotly::add_trace`, and `plotly::layout`来描述

`trace`主要描述图形上数据的相关属性

`layout`描述类似于标题、坐标轴、注释等属性

`add_*`() 类似于ggplot2中的图层

data-plot-pipeline 数据处理和图形绘制链式操作

Reference



<https://plot.ly/r>

<https://plot.ly/r/reference/>

https://cpsievert.github.io/plotly_book/

<https://plot.ly/create/>

Leaflet包是最受欢迎的开源JS交互式地图包。

功能：

- 交互式（平移、缩放）

- 多种元素组合构图：

 - Map tiles

 - Markers

 - Polygons

 - Lines

 - Popups

 - GeoJSON

- 从R或者RSutido创建地图窗口

- 可以把地图嵌入R包所生成的Markdown文档中，或者是Shiny制作的APP中

- 容易渲染由sp、sf包生成的含有经纬度的空间对象

- 可以设定地图范围以及封装自定义的鼠标事件



```
install.packages("leaflet")
library(leaflet)

m <- leaflet() %>%
  addTiles() %>% # Add default OpenStreetMap map tiles
  addMarkers(lng=123.433228, lat=41.661087, popup="东北大学")
m
```

创建leaflet map 的三个基本步骤:

- 1 调用leaflet()
- 2 使用图层函数添加图层(e.g. addTiles, addMarkers, addPolygons)
- 3 display

地图方法

`setView()`: 设定地图的显示级别缩放比例、和地图的中心点。

`fitBounds()`: 设定地图的范围，一般是一个矩形，结构是：[lng1, lat1] – [lng2, lat2]。

`clearBounds()`: 清除地图的范围设定。

数据对象

leaflet包支持各种与空间信息有关的对象，包括使用sp包定义的空间对象，和R语言中带有空间信息的数据框等，如下所示：

与R相关的：

- 由经纬度信息组成矩阵
- 带有经纬度字段的数据框。

与sp包相关的：

- `SpatialPoints[DataFrame]`
- `Line/Lines`
- `SpatialLines[DataFrame]`
- `Polygon/Polygons`
- `SpatialPolygons[DataFrame]`

基础底图

➤ Default (OpenStreetMap) Tiles

➤ Third-Party Tiles

addProviderTiles()

```
# OpenStreetMap.Mapnik
# OpenStreetMap.BlackAndWhite
# OpenStreetMap.DE
# OpenStreetMap.France
# OpenStreetMap.HOT
# OpenTopoMap
# Thunderforest.OpenCycleMap
# Thunderforest.Transport
# Thunderforest.TransportDark
# Thunderforest.SpinalMap
# Thunderforest.Landscape
# Thunderforest.Outdoors
# Thunderforest.Pioneer
```

```
# OpenMapSurfer.Roads
# OpenMapSurfer.Grayscale
# Hydda.Full
# Stamen.Toner
# Stamen.TonerBackground
# Stamen.TonerLite
# Stamen.Watercolor
# Stamen.Terrain
# Stamen.TerrainBackground
# Stamen.TopOSMRelief
# Esri.WorldStreetMap
# Esri.DeLorme
# Esri.WorldTopoMap
# Esri.WorldImagery
# Esri.WorldTerrain
```

```
# Esri.WorldShadedRelief
# Esri.WorldPhysical
# Esri.OceanBasemap
# Esri.NatGeoWorldMap
# Esri.WorldGrayCanvas
# MtbMap
# CartoDB.Positron
# CartoDB.PositronNoLabels
# CartoDB.PositronOnlyLabels
# CartoDB.DarkMatter
# CartoDB.DarkMatterNoLabels
# CartoDB.DarkMatterOnlyLabels
# HikeBike.HikeBike
# HikeBike.HillShading
# NASAGIBS.ModisTerraTrueColorCR
# NASAGIBS.ModisTerraBands367CR
# NASAGIBS.ViirsEarthAtNight2012
# NASAGIBS.ModisTerraLSTDay
# NASAGIBS.ModisTerraSnowCover
```



标记

`addMarkers()`

Customizing Marker Icons

Awesome Icons

Marker Clusters

Circle Markers

弹出窗口和几何要素

leaflet推荐学习网址：<http://rstudio.github.io/leaflet/markers.html>
和<http://blog.csdn.net/allenlu2008/article/details/52816708>



数据介绍:

systation.csv 中包含的是沈阳地铁站相关线路的经纬度数据, 字段介绍如下表:

变量名称	描述
station	站名
line	线路编号
line_id	站编号
gps_lon	站处于的经度
gps_lat	站处于的维度

SY-20150401.csv中提供的是一卡通刷卡数据, 字段介绍如下表:

变量名称	描述
V1	卡编号
V2	刷卡日期
V3	刷卡时间
V4	刷卡站点
V5	刷卡类型
V6	票价
V7	是否优惠

注: 输入数据后将字段名, 重命名为
`c('card_id','date','time','station','vehicle',
money','property')`



题目：

1. 使用leaflet绘制地铁线路图，要求

1) 图中绘制地铁线路

2) 各站点展示进站流量（08:00:00-08:05:00间的数据），流量的大小用标记的大小表示，并且提示线路、站点、流量的具体数值。

```
Shenyang <- leaflet() %>%
```

```
setView(lng=123.44,lat=41.81,zoom = 11) %>% addProviderTiles("CartoDB.Positron")
```

2.使用plotly绘制（17:00:00-17:05:00）出站流量最多的五个站点的出站流量(柱状图)。

THANKS

