IST 687 Working with map data

Corey Jackson

2021-02-24 21:26:03

Today's Agenda

- Announcements
- Exam Logistics
- Review up Week 6 - Introduction to data visualization
- Week 7 - Working with map data
  - Breakout (Lab 7)
  - Homework 7 Tips
- Next week's agenda

Announcements

- Office Hours: Wed. 6-7pm EDT and by appointment
- Upcoming Schedule
  - Week 8 (30 min. live session)
  - Complete mid-term with 72 hours *Due Saturday March 6th by 9:30pm EST*
- Project Update III in Week 10 (not Week 9)

Exam Logistics

Exam Logistics

- **Format**
  - Closed book/notes/R
  - 1 hour time limit (no pausing)
- **Materials covered**: Weeks 1-8
- **Question types**
  - Given code what is the expected output: 2
  - Write code to perform: 10
  - Open-ended questions: 9

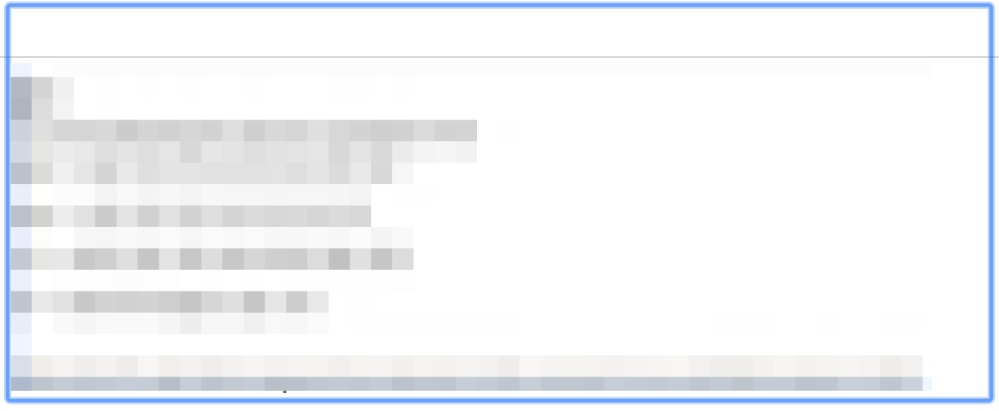
Question distribution

Week	# Questions
2 - Using R to manipulate data.	8
3 - Descriptive Statistics & Functions	5
4 - Inferential statistics	4
6 - Introduction to visualization	1
7 - Working with map data	1
8 - Linear modeling	2

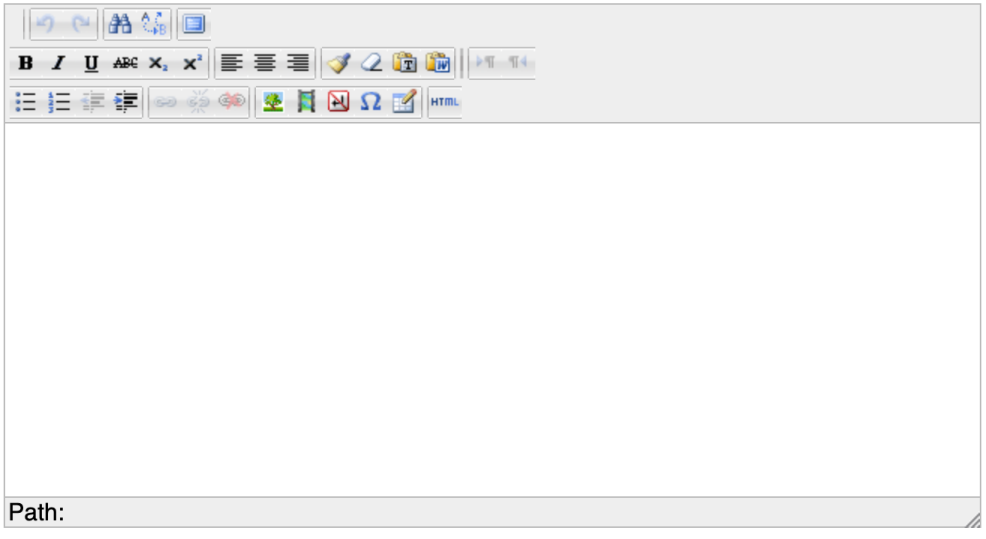
Exam office hours: Tuesday, March 2nd from 6-8pm EST or post questions on SLACK

Exam interface

Midterm Quiz



Answer:



QUIZ REPORTS

- Info
- Overview & Regrade
- Manual Grading
- Item Analysis
- **Preview**
- Quiz Settings
- User Overrides

Go to Gradebook to publish scores to students &raquo

1	2	3	4	5
6	7	8	9	10
11	12	13	14	15
16	17	18	19	20
21				

Finish Exam

Time left **0:59:45**

Start a new preview

Review: Week 6 - Introduction to data visualization

Week 6 - Introduction to visualization

- Creating visualizations with 'ggplot()'
  - Components of ggplot: data, aesthetics, geometry
- Plots for data exploration using layers: distributions `geom_histogram()`, boxplots `geom_boxplot()`, line charts `geom_line()`, heatmaps `geom_tile()`
- Adding complexity (information) by manipulating graph aesthetics e.g., `fill =`, `color =`, and `size =` and adjusting axis angles e.g., `theme(axis.text.x=element_text(angle=45, hjust=1))`

Week 6 - Additional resources on data visualization

- Practice makes perfect
  - Workshop Examples
  - Visualizing ecology data
- Good reads on visualization
  - Storytelling with Data: A Data Visualization Guide for Business Professionals by Cole Nussbaumer Knaflic
  - Choosing the right visualizations: Multiple views visualization research explained and Data Visualization 101: How to Choose the Right Chart or Graph for Your Data
- Selecting visualizations guide

Week 7 - Working with map data

Week 7 - Working with map data

- Exploring creating maps and overlaying data on maps.
  - Adding complexity to maps e.g., changing colors based on attributes of records.
- Retrieving data from open APIs (e.g., Google/OpenStreetMap)
- Using base datasets in R
- Data munging

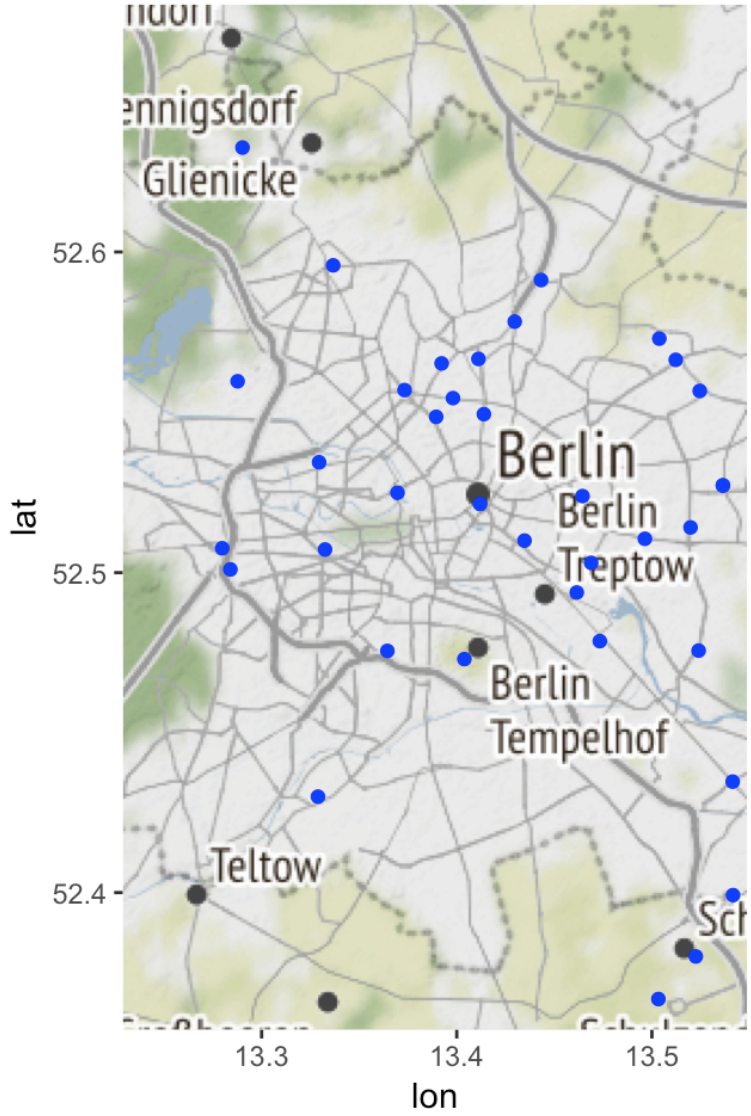
Lab 7 (50 mins.)

New Lab Assignment

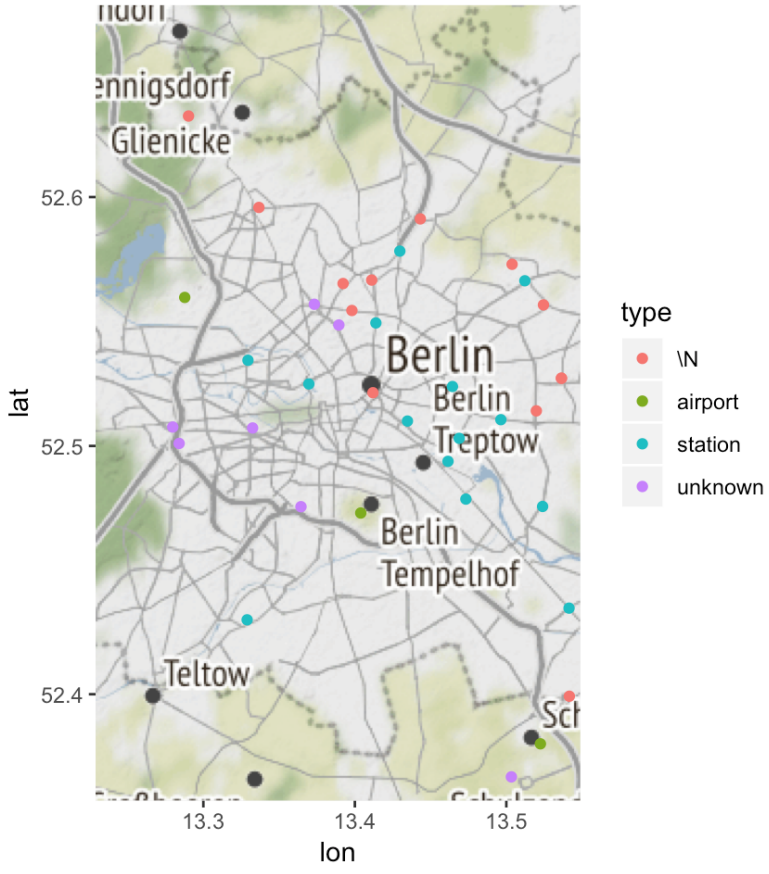
- Lab 7
- Google no longer allows API requests without a user account (its where we retrieve lat/long data for maps)
- *Objective*: Plot ports as points (e.g., airports, train stations) in a city of your choosing using OpenStreetMap data
- *Dataset*: World Ports

Lab 7 Output

- Your output should look similar to these maps of the ports in Berlin



Lab 7 Output


Glienicke
52.6
52.5
52.4
lat
Berlin
Berlin
Treptow
Berlin
Tempelhof
Teltow
13.3
13.4
13.5
lon
type
\N
airport
station
unknown

Lab 7 Packages needed to complete lab

```
library(ggplot2) # To plot map aesthetics
library(ggmap) # Get city map
library(readr) # reads in the OSM data
library(tmaptools) # For the geocode_OSM function
  in "Getting a base map"
```

Week 7 Homework Tips

Week 7 Homework tips

Goal: Practice creating maps and overlaying data - Data munging necessary to obtain information about U.S. states

- Load zipcode data (Step 1, question 3).
- Install the `zipcode()` package and load the zipcode data using `data(zipcode)`

```
##     zip        city state latitude longitude
## 1 00210 Portsmouth    NH  43.0059  -71.0132
## 2 00211 Portsmouth    NH  43.0059  -71.0132
## 3 00212 Portsmouth    NH  43.0059  -71.0132
## 4 00213 Portsmouth    NH  43.0059  -71.0132
```

Week 7 Homework tips: A manual install of the zipcodes package

- The CRAN download of the `zipcode()` package is no longer available.
- Archived versions of the package are made available: [`zipcode()`] (http://www.coreybjackson.com/687/Datasets/zipcode_1.0.tar.gz).

To install the package follow the steps below:

1. Download the package files to your desktop.
2. Navigate to where the files downloaded and unzip the file. You should be able to simply click the zip file and its contents will unload in the same directory, but in a folder.
3. Navigate back to RStudio and install using this code: install.packages("~/Downloads/zipcode/", repos = NULL, type="source") You'll want to set the first argument to the location of the file. In my example "~/Downloads/zipcode/" is the location of the file I unzipped in step 2.

Week 7 Homework Tips: Cleaning State Data

- Cleaning zip codes using `clean.zipcodes()` (Step 1, 4).

```
## [1] "2061"       "02142"      "2043"       "20210"
## [6] "SW1P 3JX"   "210"        "02199-1880"
```

```
clean.zipcodes(zips)
```

```
## [1] "02061" "02142" "02043" "20210" "02061" NA      "002
```

- In HW, clean zip codes in your data before merging

Week 7 Homework Tips: Cleaning State Data

- Retrieve the names of states and abbreviations of states (Step 2, question 2). Simply call `state.name` and `state.abb`. You'll need to use these vectors adding them to a dataframe.

```
state.name
```

```
## [1] "Alabama"    "Alaska"     "Arizona"    "Arkansas"
## [6] "Colorado"
```

```
state.abb
```

```
## [1] "AL" "AK" "AZ" "AR" "CA" "CO"
```

Week 7 Homework Tips: Merging dataframes

- Merging dataframes using `merge()` (Step 2)
  - Basic structure merge two data frames by one ID: `merge(x,y, by= "ID", ...)`
  - Example dataframes: `authors` and `books`

```
##    surname nationality retired
## 1    Tukey          US     yes
## 2 Venables   Australia      no
## 3  Tierney          US      no
## 4   Ripley          UK      no
```

```
##       name                              title other.author
## 1    Tukey     Exploratory Data Analysis          <NA>
## 2 Venables Modern Applied Statistics ...        Ripley
## 3  Tierney                  LISP-STAT          <NA>
## 4   Ripley            Spatial Statistics          <NA>
```

Week 7 Homework Tips: Merging dataframes

- Each dataframe needs a common column on which to merge
- In the authors and books data `surname` and `name` are common columns

```
merge(authors, books, by.x="surname", by.y="name")
```

```
##    surname nationality retired                              t
## 1   McNeil   Australia      no     Interactive Data Anal
## 2   Ripley          UK      no           Spatial Statis
## 3   Ripley          UK      no       Stochastic Simula
## 4  Tierney          US      no                    LISP-
## 5    Tukey          US     yes     Exploratory Data Anal
## 6 Venables   Australia      no Modern Applied Statistics
```

Next week

- **Asynchronous Materials**
  - Week 8: Linear regression (be sure to cover for exam)
  - Submit HW 7 & Lab 7 Monday
  - Continue working on group project (Project Update 3 due in Week 10)
- **Synchronous Session**
  - Overview of Data Mining and Week 8 (30 minutes)
  - Mid-term questions (Mid-term to be completed within 72 hours)