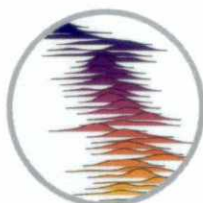
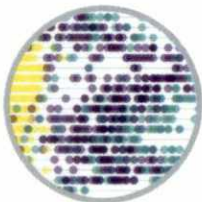
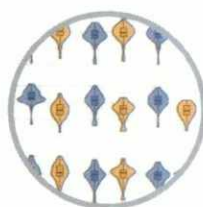
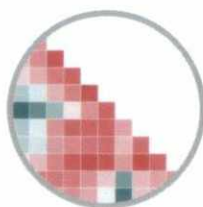
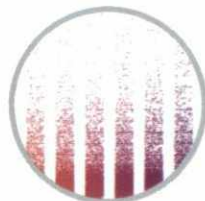
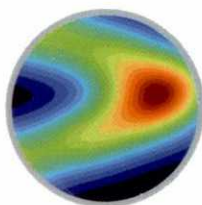


Data Visualization in **R** and **Python**

MARCO CREMONINI




WILEY

Contents

Preface *xiii*

Introduction *xv*

About the Companion Website *xxiii*

Part I Static Graphics with ggplot (R) and Seaborn (Python) *1*

1 Scatterplots and Line Plots *3*

1.1 R: ggplot *4*

1.1.1 Scatterplot *4*

1.1.2 Repulsive Textual Annotations: Package *ggrepel* *13*

1.1.3 Scatterplots with High Number of Data Points *15*

1.1.4 Line Plot *17*

1.2 Python: Seaborn *19*

1.2.1 Scatterplot *21*

1.2.2 Line Plot *25*

2 Bar Plots *29*

2.1 R: ggplot *29*

2.1.1 Bar Plot and Continuous Variables: Ranges of Values *33*

2.2 Python: Seaborn *34*

2.2.1 Bar Plot with Three Variables *35*

2.2.2 Ranges of Values from a Continuous Variable *37*

2.2.3 Visualizing Subplots *39*

3 Facets *43*

3.1 R: ggplot *44*

3.1.1 Case 1: Temperature *44*

3.1.2	Case 2: Air Quality	45
3.2	Python: Seaborn	49
3.2.1	Facet for Scatterplots and Line Plot	50
3.2.2	Line Plot	50
3.2.3	Facet and Graphics for Categorical Variables	51
3.2.4	Facet and Bar Plots	51
3.2.5	Facets: General Method	54
4	Histograms and Kernel Density Plots	59
4.1	R: ggplot	59
4.1.1	Univariate Analysis	60
4.1.2	Bivariate Analysis	63
4.1.3	Kernel Density Plots	67
4.2	Python: Seaborn	71
4.2.1	Univariate Analysis	71
4.2.2	Bivariate Analysis	73
4.2.3	Logarithmic Scale	75
5	Diverging Bar Plots and Lollipop Plots	83
5.1	R: ggplot	83
5.1.1	Diverging Bar Plot	83
5.1.2	Lollipop Plot	89
5.2	Python: Seaborn	91
5.2.1	Diverging Bar Plot	91
6	Boxplots	99
6.1	R: ggplot	100
6.2	Python: Seaborn	105
7	Violin Plots	109
7.1	R: ggplot	110
7.1.1	Violin Plot and Scatterplot	113
7.1.2	Violin Plot and Boxplot	114
7.2	Python: Seaborn	117
8	Overplotting, Jitter, and Sina Plots	121
8.1	Overplotting	121
8.2	R: ggplot	122
8.2.1	Categorical Scatterplot	122
8.2.2	Violin Plot and Scatterplot with Jitter	123
8.2.3	Sina Plot	126

8.2.4	Beeswarm Plot	129
8.2.5	Comparison Between Jittering, Sina plot, and Beeswarm plot	131
8.3	Python: Seaborn	131
8.3.1	Strip Plot and Swarm Plot	131
8.3.2	Sina Plot	134
9	Half-Violin Plots	137
9.1	R: ggplot	138
9.1.1	Custom Function	138
9.1.2	Raincloud Plot	141
9.2	Python: Seaborn	144
10	Ridgeline Plots	147
10.1	History of the Ridgeline	147
10.2	R: ggplot	148
11	Heatmaps	157
11.1	R: ggplot	157
11.2	Python: Seaborn	160
12	Marginals and Plots Alignment	165
12.1	R: ggplot	165
12.1.1	Marginal	165
12.1.2	Plots Alignment	166
12.1.3	Rug Plot	168
12.2	Python: Seaborn	170
12.2.1	Subplots	170
12.2.2	Marginals: Joint Plot	173
12.2.3	Marginals: Joint Grid	173
13	Correlation Graphics and Cluster Maps	177
13.1	R: ggplot	178
13.1.1	Cluster Map	178
13.2	Python: Seaborn	182
13.2.1	Cluster Map	182
13.3	R: ggplot	184
13.3.1	Correlation Matrix	184
13.4	Python: Seaborn	184
13.4.1	Correlation Matrix	184
13.4.2	Diagonal Correlation Heatmap	186
13.4.3	Scatterplot Heatmap	188

Part II Interactive Graphics with Altair 193**14 Altair Interactive Plots 195**

- 14.1 Scatterplots 196
 - 14.1.1 Static Graphics 197
 - 14.1.1.1 JSON Format: Data Organization 200
 - 14.1.1.2 Plot Alignment and Variable Types 201
 - 14.1.2 Facets 202
 - 14.1.3 Interactive Graphics 205
 - 14.1.3.1 Dynamic Tooltips 205
 - 14.1.3.2 Interactive Legend 207
 - 14.1.3.3 Dynamic Zoom 208
 - 14.1.3.4 Mouse Hovering and Contextual Change of Color 210
 - 14.1.3.5 Drop-Down Menu and Radio Buttons 212
 - 14.1.3.6 Selection with Brush 214
 - 14.1.3.7 Graphics as Legends 220
- 14.2 Line Plots 225
 - 14.2.1 Static Graphics 225
 - 14.2.2 Interactive Graphics 228
 - 14.2.2.1 Highlighted Lines with Mouse Hover 228
 - 14.2.2.2 Aligned Tooltips 231
- 14.3 Bar Plots 235
 - 14.3.1 Static Graphics 235
 - 14.3.1.1 Diverging Bar Plot 239
 - 14.3.1.2 Plots with Double Scale 240
 - 14.3.1.3 Stacked Bar Plots 244
 - 14.3.1.4 Sorted Bars 246
 - 14.3.2 Interactive Graphics 247
 - 14.3.2.1 Synchronized Bar Plots 247
 - 14.3.2.2 Bar Plot with Slider 251
- 14.4 Bubble Plots 257
 - 14.4.1 Interactive Graphics 257
 - 14.4.1.1 Bubble Plot with Slider 257
- 14.5 Heatmaps and Histograms 260
 - 14.5.1 Interactive Graphics 260
 - 14.5.1.1 Heatmaps 260
 - 14.5.1.2 Histograms 262

Part III Web Dashboards 267**15 Shiny Dashboards 271**

- 15.1 General Organization 271

- 15.2 Second Version: Graphics and Style Options 280
- 15.3 Third Version: Tabs, Widgets, and Advanced Themes 286
- 15.4 Observe and Reactive 289

16 Advanced Shiny Dashboards 295

- 16.1 First Version: Sidebar, Widgets, Customized Themes, and Reactive/Observe 295
 - 16.1.1 Button Widget: Observe Context 297
 - 16.1.2 Button Widget: Mode of Operation 298
 - 16.1.3 HTML Data Table 301
- 16.2 Second Version: Tabs, *Shinydashboard*, and Web Scraping 303
 - 16.2.1 Shiny Dashboard 303
 - 16.2.2 Web Scraping of HTML Tables 308
 - 16.2.3 Shiny Dashboards and Altair Graphics Integration 315
 - 16.2.4 Altair and *Reticulate*: Installation and Configuration 319
 - 16.2.5 Simple Dashboard for Testing Shiny-Altair Integration 320
- 16.3 Third Version: Altair Graphics 321
 - 16.3.1 Cleveland Plot and Other Graphics 325

17 Plotly Graphics 329

- 17.1 Plotly Graphics 329
 - 17.1.1 Scatterplot 331
 - 17.1.2 Line Plot 334
 - 17.1.3 Marginals 334
 - 17.1.4 Facets 334

18 Dash Dashboards 339

- 18.1 Preliminary Operations: Import and Data Wrangling 340
 - 18.1.1 Import of Modules and Submodules 340
 - 18.1.2 Data Import and Data-Wrangling Operations 341
- 18.2 First Dash Dashboard: Base Elements and Layout Organization 341
 - 18.2.1 Plotly Graphic 341
 - 18.2.2 Themes and Widgets 342
 - 18.2.3 Reactive Events and Callbacks 344
 - 18.2.4 Data Table 345
 - 18.2.5 Color Palette Selector and Data Table Layout Organization 348
- 18.3 Second Dash Dashboard: Sidebar, Widgets, Themes, and Style Options 355
 - 18.3.1 Sidebar, Multiple Selection, and Checkbox 355
 - 18.3.2 Dark Themes 360
 - 18.3.3 Radio Buttons 361
 - 18.3.4 Bar Plot 363
 - 18.3.5 Container 364

18.4	Third Dash Dashboard: Tabs and Web Scraping of HTML Tables	365
18.4.1	Multi-page Organization: Tabs	365
18.4.2	Web Scraping of HTML Tables	370
18.4.3	Second Tab's Layout	371
18.4.4	Second Tab's Reactive Events	372
18.5	Fourth Dash Dashboard: Light Theme, Custom CSS Style Sheet, and Interactive Altair Graphics	377
18.5.1	Light Theme and External CSS Style Sheet	377
18.5.2	Altair Graphics	379

Part IV Spatial Data and Geographic Maps 389

19	Geographic Maps with R	391
19.1	Spatial Data	392
19.2	Choropleth Maps	397
19.2.1	Eurostat – GISCO: giscoR	400
19.3	Multiple and Annotated Maps	404
19.3.1	From ggplot to Plotly Graphics	408
19.4	Spatial Data (sp) and Simple Features (sf)	408
19.4.1	Natural Earth	408
19.4.2	Format sp and sf: Centroid and Polygons	410
19.4.3	Differences Between Format sp and Format sf	411
19.5	Overlaid Graphical Layers	413
19.6	Shape Files and GeoJSON Datasets	419
19.7	Venice: Open Data Cartography and Other Maps	420
19.7.1	Tiled Web Maps	430
19.7.1.1	Package ggmap	430
19.7.1.2	Package Leaflet	431
19.7.2	Tiled Web Maps and Layers of sf Objects	433
19.7.2.1	Tiled Web Maps with ggmap	435
19.7.2.2	Tiled Web Map with Leaflet	440
19.7.3	Maps with Markers and Annotations	445
19.8	Thematic Maps with tmap	448
19.8.1	Static and Interactive Visualizations	451
19.8.2	Cartographic Layers: Rome's Archaeological Sites	457
19.9	Rome's Accommodations: Intersecting Geometries with Simple Features and tmap	460
19.9.1	Centroids and Active Geometry	462
19.9.2	Quantiles and Custom Legend	466
19.9.3	Variants with Points and Popups	473

20	Geographic Maps with Python	481
20.1	New York City: Plotly	481
20.1.1	Choropleth Maps: <i>plotly.express</i>	484
20.1.1.1	Dynamic Tooltips	485
20.1.1.2	Mapbox	487
20.1.2	Choropleth Maps: <i>plotly.graph_objects</i> (<i>plotly go</i>)	489
20.1.3	GeoJSON Polygon, Multipolygon, and Missing <i>id</i> Element	490
20.2	Overlaid Layers	491
20.3	Geopandas: Base Map, Data Frame, and Overlaid Layers	495
20.3.1	Extended Dynamic Tooltips	496
20.3.2	Overlaid Layers: Dog Breeds, Dog Runs, and Parks Drinking Fountains	500
20.4	Folium	507
20.4.1	Base Maps, Markers, and Circles	508
20.4.2	Advanced Tooltips and Popups	511
20.4.3	Overlaid Layers and GeoJSON Datasets	514
20.4.4	Choropleth Maps	515
20.4.5	Geopandas	518
20.4.6	Folium Heatmap	520
20.5	Altair: Choropleth Map	522
20.5.1	GeoJSON Maps	523
20.5.2	Geopandas: NYC Subway Stations and Demographic Data	523

Index 529