

Visualization of Sequence Data

Sonali Arora*

27-28 February 2014

Contents

1 Visualization

1

1 Visualization

R has some great visualization packages; Here we take a quick tour of visualization facilities tailed for sequence data and using Bioconductor approach using *Gviz*. The *Gviz* package produces very elegant data organized in a more-or-less familiar 'track' format. The following exercises walk through the Gviz User guide Section 2.

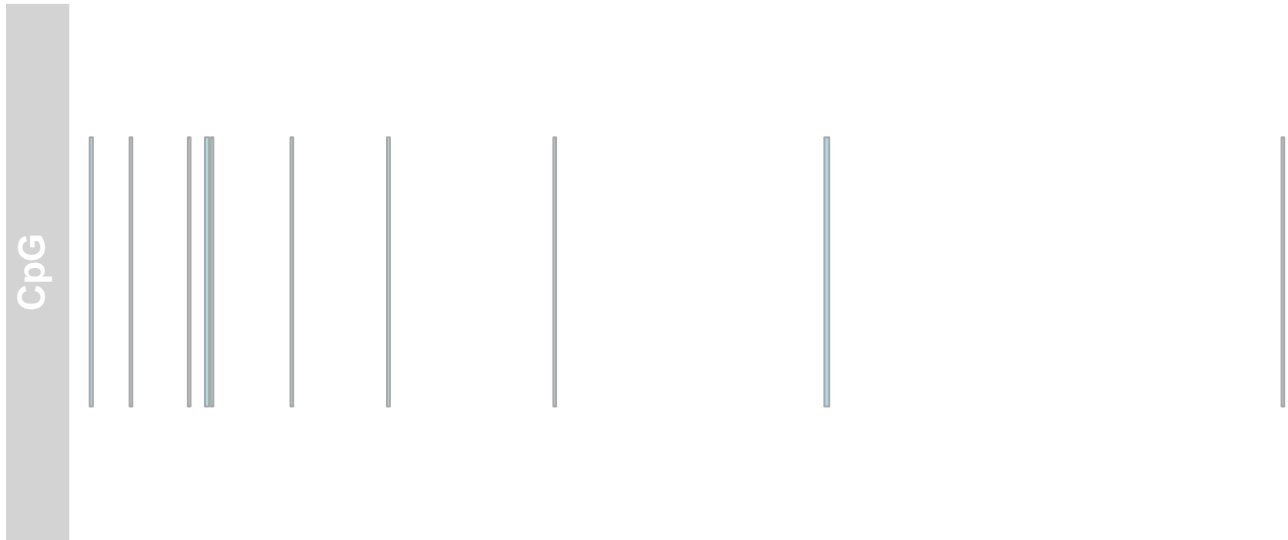
Exercise : *Gviz* visualization Load the Gviz package and sample GRanges containing genomic coordinates of CpG islands. Create a couple of variables with information on the chromosome and genome of the data (how can this information be extracted from the cpGIslands object?).

```
library(Gviz)
data(cpGIslands)
chr <- "chr7"
genome <- "hg19"
```

The basic idea is to create a track, perhaps with additional attributes, and to plot it. There are different types of track, and we create these one at a time. We start with a simple annotation track

```
atrack <- AnnotationTrack(cpGIslands, name="CpG")
plotTracks(atrack)
```

*sarora@fhcrc.org



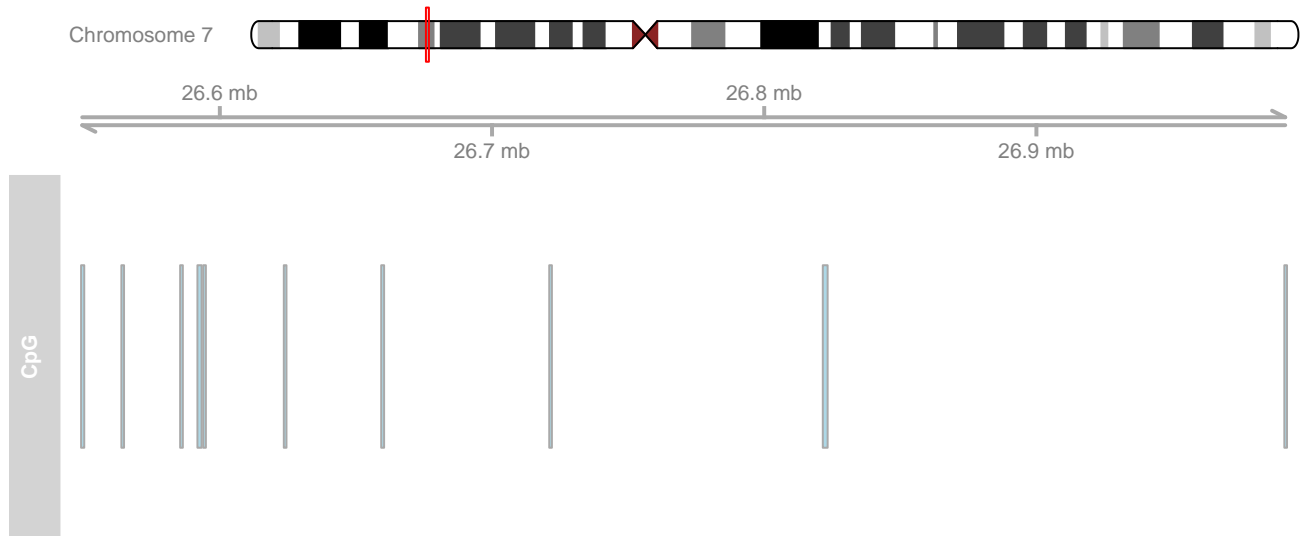
Then add a track that represents genomic coordinates. Tracks are combined during when plotted, as a simple list. The vertical ordering of tracks is determined by their position in the list.

```
gtrack <- GenomeAxisTrack()
plotTracks(list(gtrack, atrack))
```



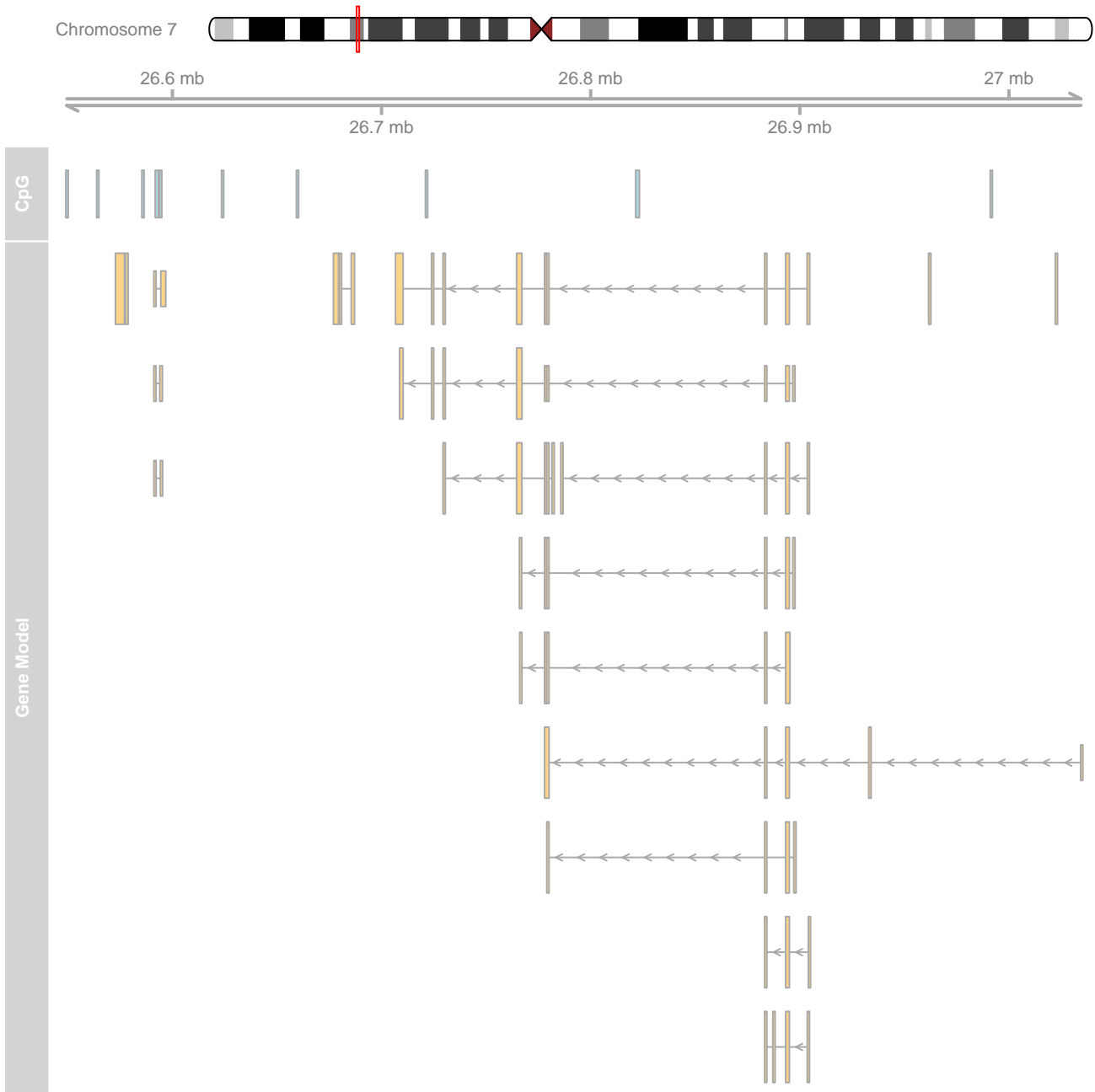
We can add an ideogram to provide overall orientation. . .

```
itrack <- IdeogramTrack(genome=genome, chromosome=chr)
plotTracks(list(itrack, gtrack, atrack))
```



And a more elaborate gene model, as an `data.frame` or `GRanges` object with specific columns of metadata.

```
data(geneModels)
grtrack <-
  GeneRegionTrack(geneModels, genome=genome,
    chromosome=chr, name="Gene Model")
tracks <- list(itrack, gtrack, atrack, grtrack)
plotTracks(tracks)
```



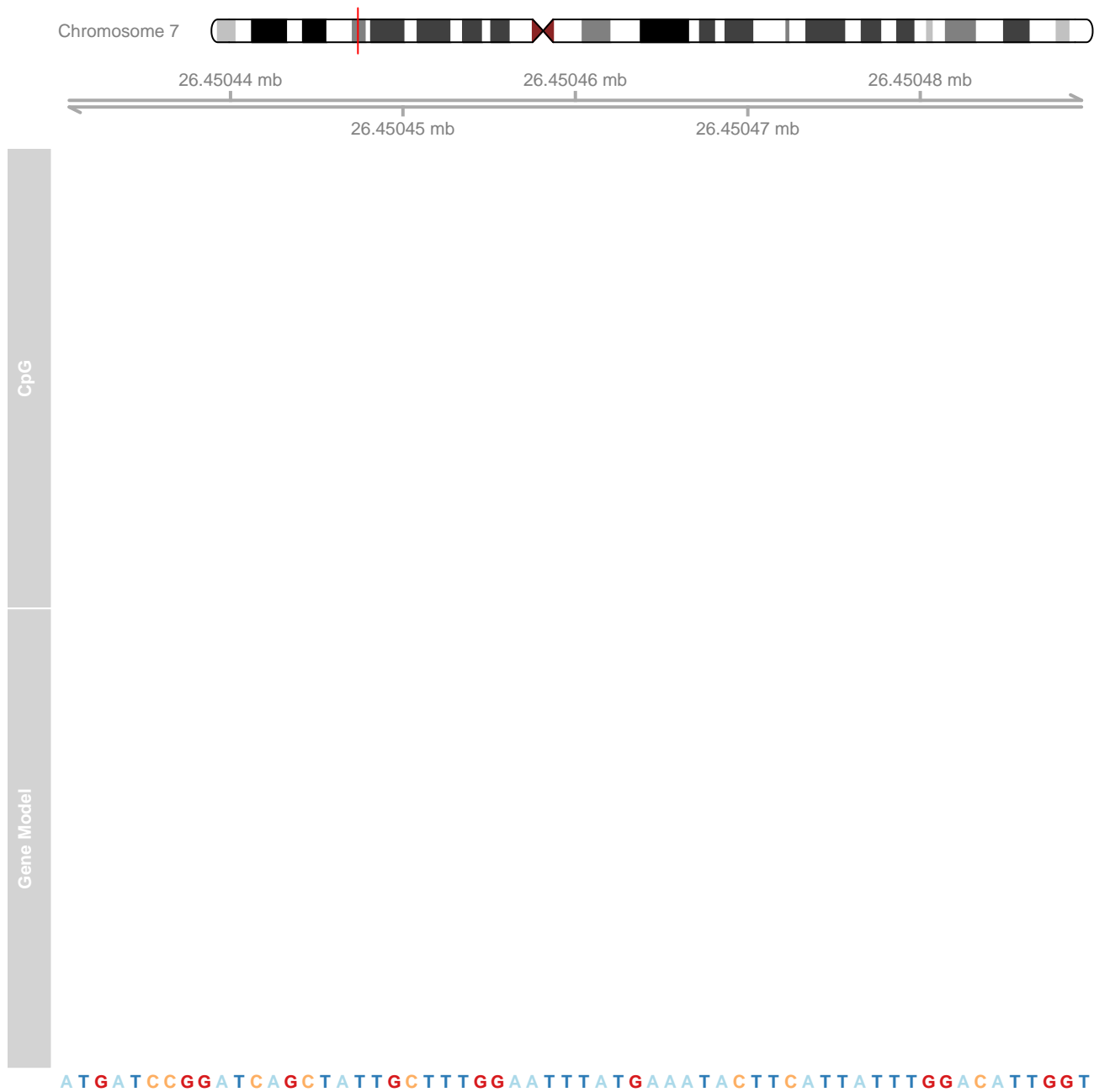
Zooming out changes the location box on the ideogram

```
plotTracks(tracks, from=2.5e7, to=2.8e7)
```



When zoomed in we can add sequence data

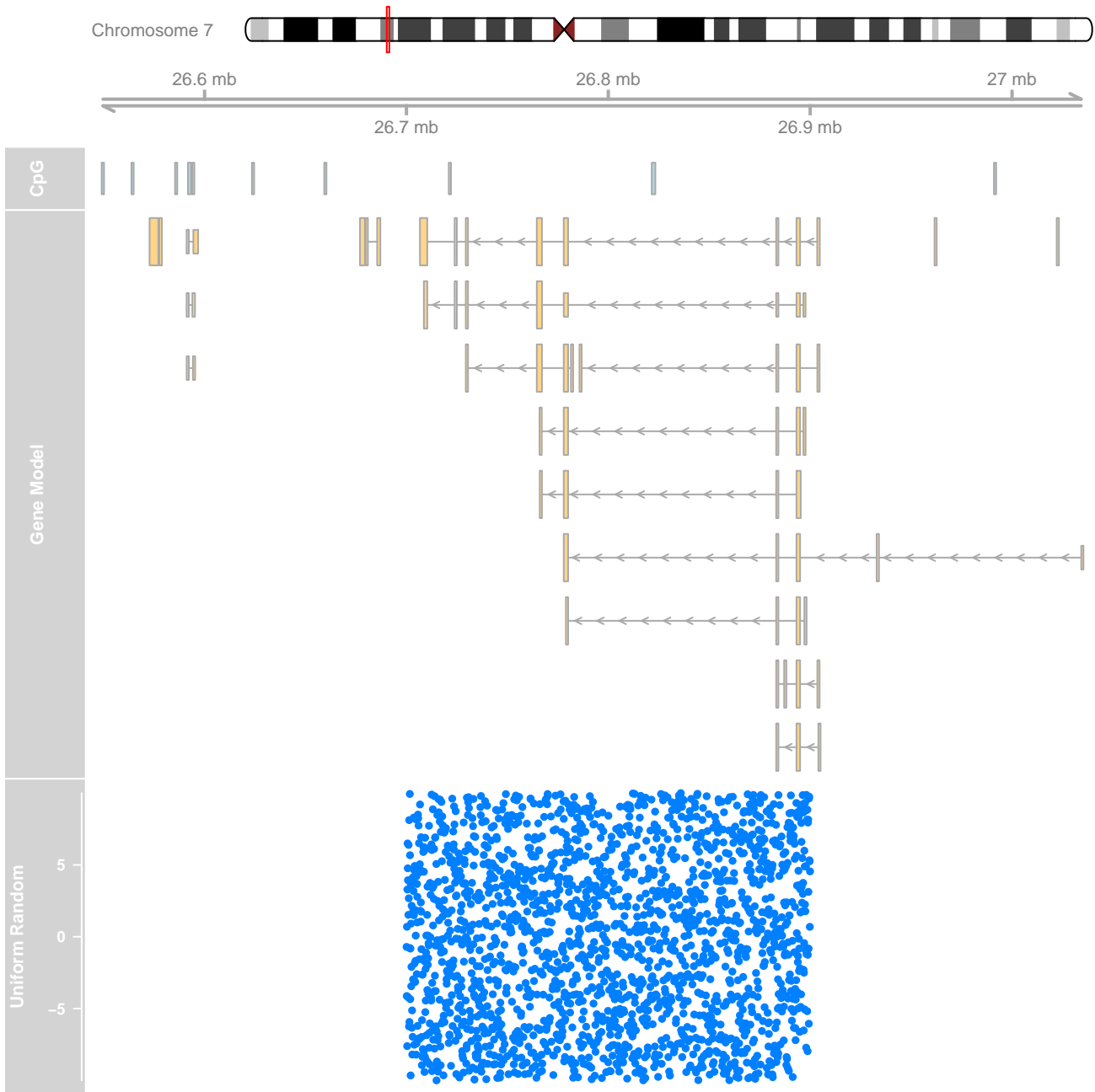
```
library(BSgenome.Hsapiens.UCSC.hg19)
strack <- SequenceTrack(Hsapiens, chromosome=chr)
plotTracks(c(tracks, strack), from=26450430, to=26450490, cex=.8)
```



As the Gviz vignette humbly says, ‘so far we have replicated the features of a whole bunch of other genome browser tools out there’. We’d like to be able integrate our data into these plots, with a rich range of plotting options. The key is the `DataTrack` function, which we demonstrate with some simulated data

```
## some data
lim <- c(26700000, 26900000)
coords <- seq(lim[1], lim[2], 101)
dat <- runif(length(coords) - 1, min=-10, max=10)
## DataTrack
dtrack <-
  DataTrack(data=dat, start=coords[-length(coords)],
            end= coords[-1], chromosome=chr, genome=genome,
```

```
name="Uniform Random")  
plotTracks(c(tracks, dtrack))
```



Section 4.3 of the Gviz vignette illustrates exibility of the data track.