Plots

Data	• Variables from the dataset
Aesthetics	• x-axis, y-axis, color, fill, size, labels, line width, line type, alpha, shape
Geometries	• Point, line, histogram, bar, boxplot
Facets	• Columns and rows
Statistics	• Binning, smoothing, descriptive stats, inferential stats
Coordinates	• Cartesian, fixed, polar, limits
Themes	• Describes the design elements. Non-data Ink.

Variable Type	`ggplot2` Function/Technique
Discrete (Categorical)	Bar Chart: `ggplot(data, aes(x=discrete_var)) + geom_bar()`
Numeric	Histogram: `ggplot(data, aes(x=numeric_var)) + geom_histogram()` Density Plot: `ggplot(data, aes(x=numeric_var)) + geom_density()`

Variable Type	`ggplot2` Function/Technique
Discrete (Categorical)	Bar Chart: `ggplot(data, aes(x=discrete_var)) + geom_bar()`
Numeric	Histogram: `ggplot(data, aes(x=numeric_var)) + geom_histogram()` Density Plot: `ggplot(data, aes(x=numeric_var)) + geom_density()`

X Variable	Y Variable	Type of Plot	`ggplot2` Function/Technique
Discrete	Discrete	Mosaic plot, Grouped bar chart	`geom_tile()`, `geom_bar(position = "dodge")`
Discrete	Numeric	Box plot, Violin plot, Jitter plot, Bar chart	`geom_boxplot()`, `geom_violin()`, `geom_jitter()`, `geom_bar(stat="summary")`
Numeric	Discrete	The above, but with axes swapped	Similar functions, but swap x and y
Numeric	Numeric	Scatter plot, Hexbin plot, 2D Density plot	`geom_point()`, `geom_hex()`, `geom_density_2d()`

Variable Type	`ggplot2` Function/Technique
Discrete (Categorical)	Bar Chart: `ggplot(data, aes(x=discrete_var)) + geom_bar()`
Numeric	Histogram: `ggplot(data, aes(x=numeric_var)) + geom_histogram()` Density Plot: `ggplot(data, aes(x=numeric_var)) + geom_density()`

X Variable	Y Variable	Type of Plot	`ggplot2` Function/Technique
Discrete	Discrete	Mosaic plot, Grouped bar chart	`geom_tile()`, `geom_bar(position = "dodge")`
Discrete	Numeric	Box plot, Violin plot, Jitter plot, Bar chart	`geom_boxplot()`, `geom_violin()`, `geom_jitter()`, `geom_bar(stat="summary")`
Numeric	Discrete	The above, but with axes swapped	Similar functions, but swap x and y
Numeric	Numeric	Scatter plot, Hexbin plot, 2D Density plot	`geom_point()`, `geom_hex()`, `geom_density_2d()`

X Variable	Y Variable	Additional Variable	Type of Plot	`ggplot2` Function/Technique
Discrete	Numeric	Numeric/Discrete	Faceted box/jitter plots, Grouped bar chart	`facet_wrap(~ additional_var) + geom_boxplot()`, `facet_wrap(~ additional_var) + geom_jitter()`, `geom_bar(position = "dodge")`
Numeric	Numeric	Discrete	Colored scatter plots, Faceted scatter plots	`geom_point(aes(color = additional_var))`, `facet_wrap(~ additional_var) + geom_point()`
Numeric	Numeric	Numeric	Bubble chart	`geom_point(aes(size = additional_var))`
Discrete	Discrete	Numeric/Discrete	Grouped bar chart with color/facets	`geom_bar(position = "dodge", aes(fill = additional_var))`, `facet_wrap(~ additional_var) + geom_bar(position = "dodge")`

Qualitative vs. quantitative

	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
Mazda RX4	21.0	6	160.0	110	3.90	2.620	16.46	0	1	4	4
Mazda RX4 Wag	21.0	6	160.0	110	3.90	2.875	17.02	0	1	4	4
Datsun 710	22.8	4	108.0	93	3.85	2.320	18.61	1	1	4	1
Hornet 4 Drive	21.4	6	258.0	110	3.08	3.215	19.44	1	0	3	1
Hornet Sportabout	18.7	8	360.0	175	3.15	3.440	17.02	0	0	3	2
Valiant	18.1	6	225.0	105	2.76	3.460	20.22	1	0	3	1
Duster 360	14.3	8	360.0	245	3.21	3.570	15.84	0	0	3	4
Merc 240D	24.4	4	146.7	62	3.69	3.190	20.00	1	0	4	2
Merc 230	22.8	4	140.8	95	3.92	3.150	22.90	1	0	4	2
Merc 280	19.2	6	167.6	123	3.92	3.440	18.30	1	0	4	4

Quantitative data

Scatter plots are used when the graph is set up to display the relationship between **two numeric** variables.
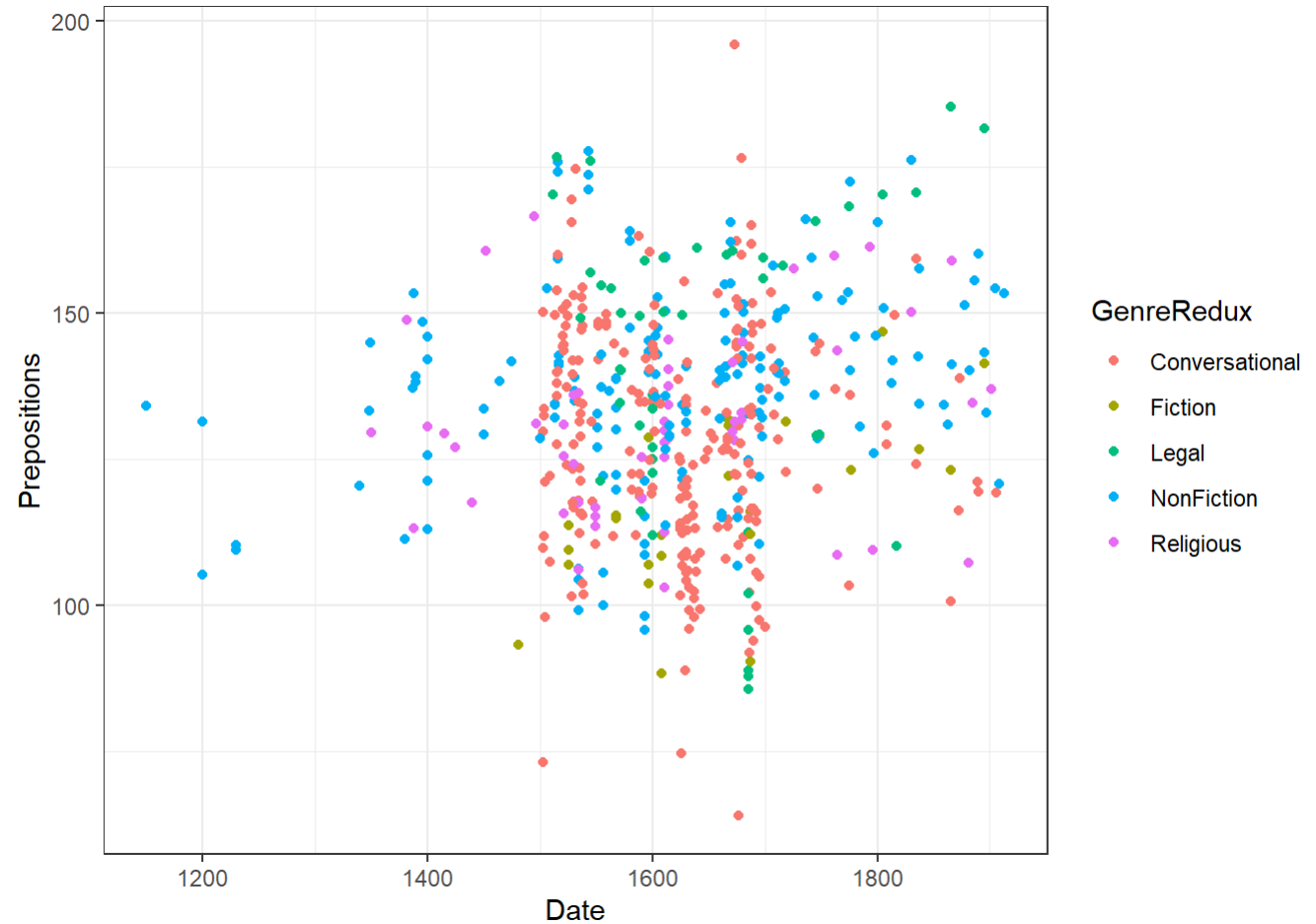

```
ggplot(subsubastra3,aes(x=headshot_percentage,y=opening_kill_rating,color=first_kill_in_won_rounds )) +
  geom_point() +
  theme_bw()
```

Distribution of the data with respect to **numeric** variables
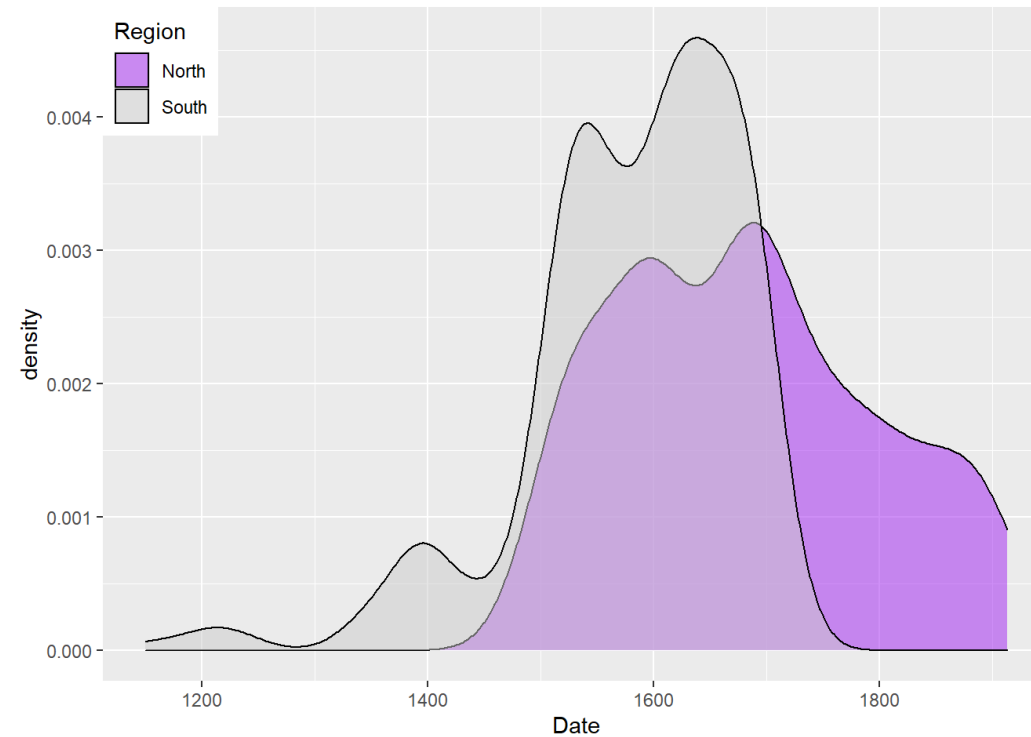

```
ggplot(pdat, aes(x = Date, y = Prepositions, color=Region)) +
geom_point() +
scale_color_manual(values = clrs[1:2]) +
theme(legend.position=c(0,1), legend.justification=c(0,1))
```

Line graphs are used when we have **numeric values** that are linked (in one way or another) because they come from the same speaker or genre as in our case).
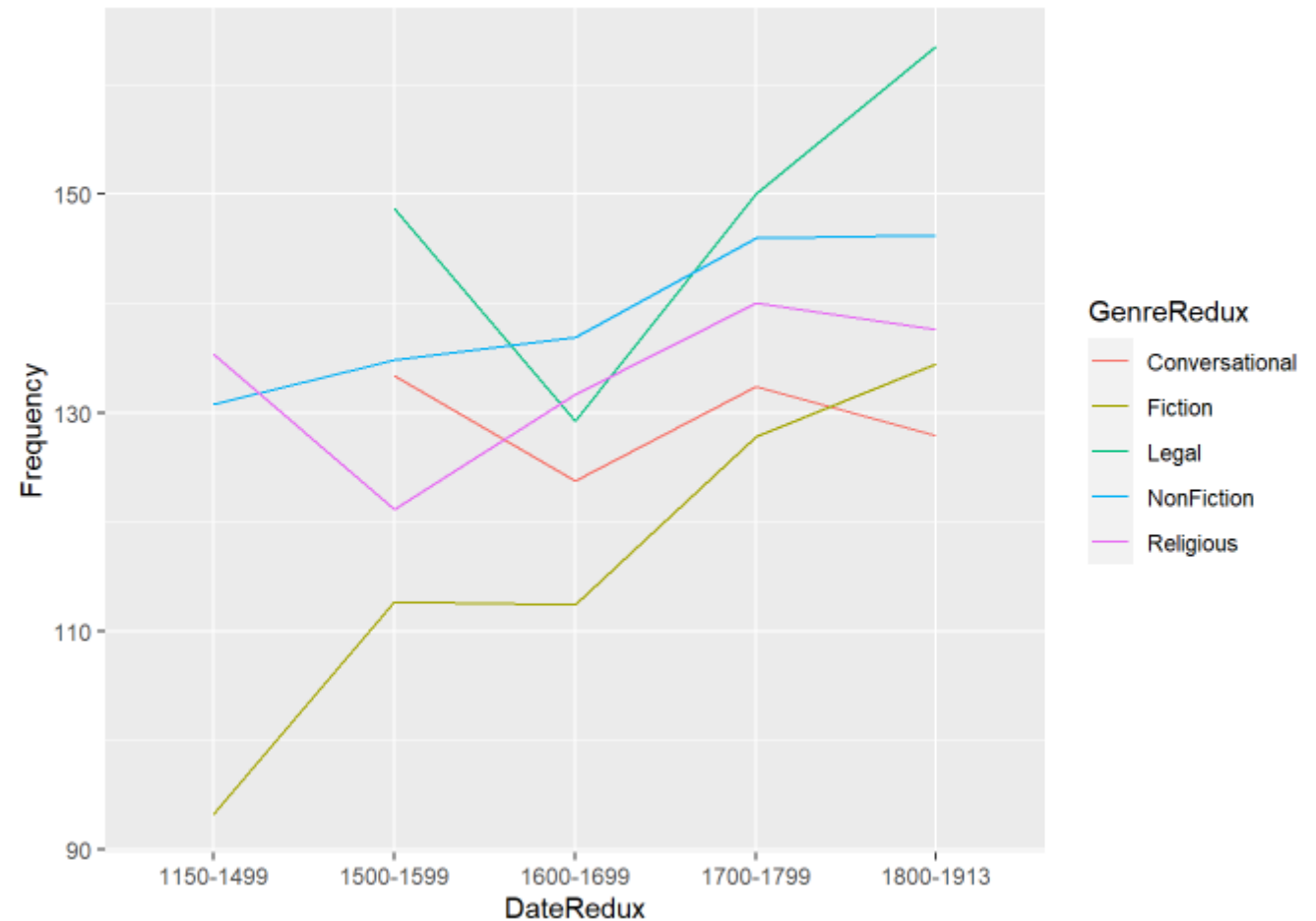
ggplot(aes(x=DateRedux, y= Frequency, group= GenreRedux, color = GenreRedux)) + geom_line()

The default way to visualize **tables of absolute frequencies** associated with a **categorical or nominal** variable are pie charts

First 15 rows of the bdat data.

DateRedux	Frequency	Percent
1150-1499	34	6.3
1500-1599	180	33.5
1600-1699	225	41.9
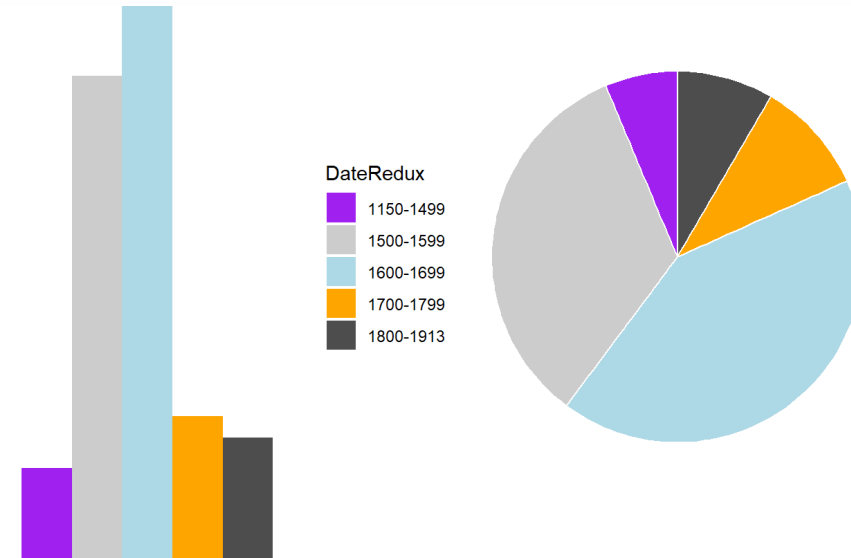

```
p1 <- ggplot(bdat, aes("", Percent, fill = DateRedux)) + geom_bar(stat="identity", position = position_dodge()) +
scale_fill_manual(values = clrs) + theme_void()

p2 <- ggplot(bdat, aes("", Percent, fill = DateRedux)) + geom_bar(stat="identity", width=1, color = "white") +
coord_polar("y", start=0) + scale_fill_manual(values = clrs) + theme_void() + theme(legend.position = "none")
grid.arrange(p1, p2, nrow = 1)
```

Histograms summarize **numeric** variables by showing their distribution across bins. We can simply add information about **a second variable** by specifying this variable as the basis for the coloring of the bars (which we do by specify the fill argument).
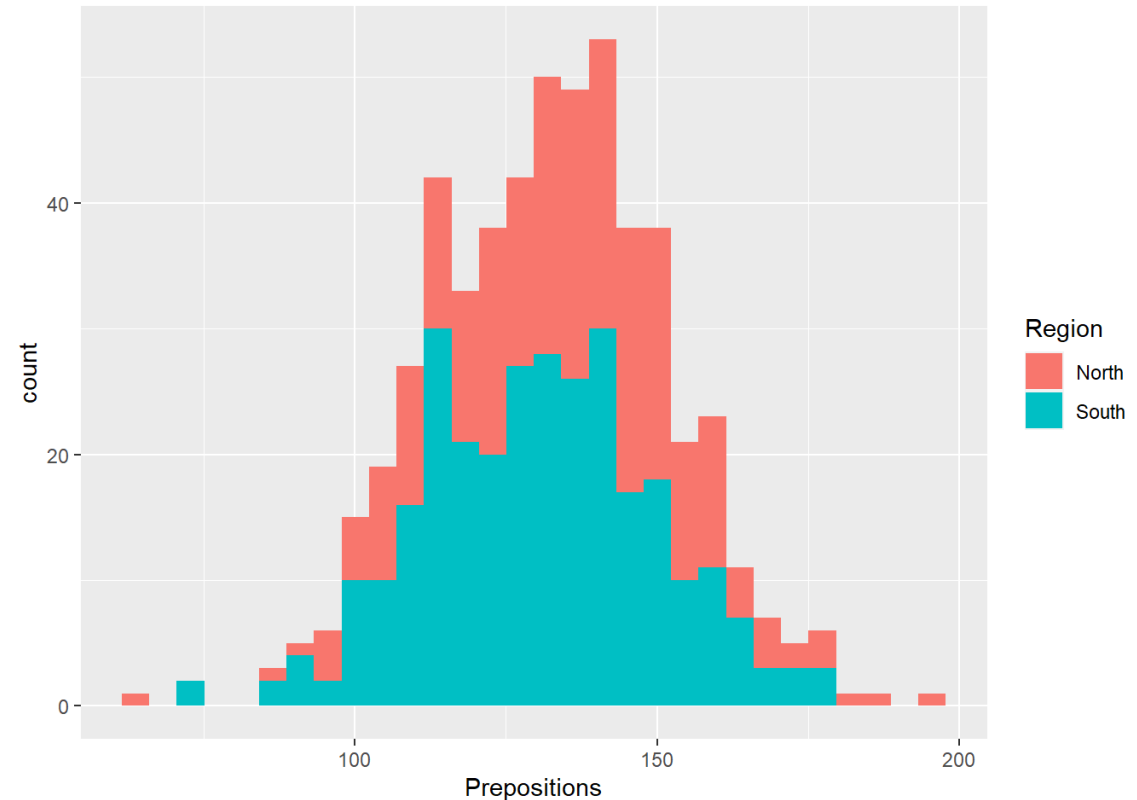

```
ggplot(pdat, aes(Prepositions, fill = Region)) +
  geom_histogram()
```

bar plot display frequency information across **categorical** variable levels.
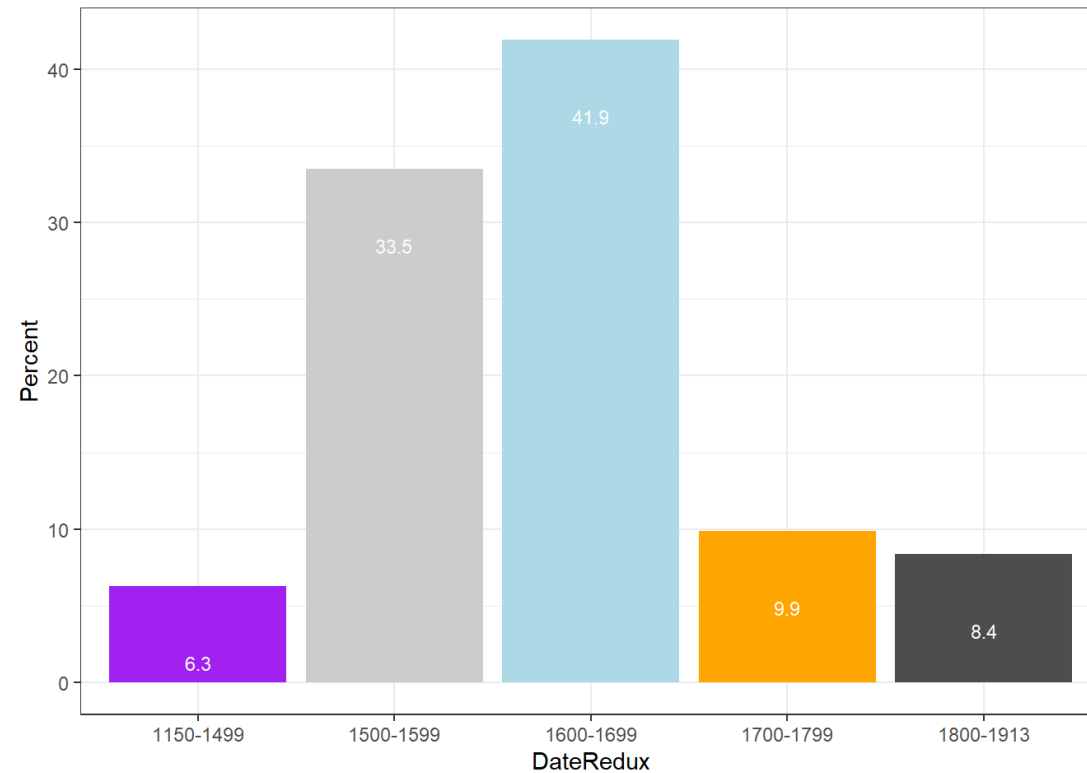
ggplot(bdat, aes(DateRedux, Percent, fill = DateRedux)) +
geom_bar(stat="identity") + theme_bw() +
geom_text(aes(y = Percent-5, label = Percent), color = "white", size=3) +
scale_fill_manual(values = clrs)

To create **grouped** bar plots, we plot Region while including DateRedux as **the fill argument**.
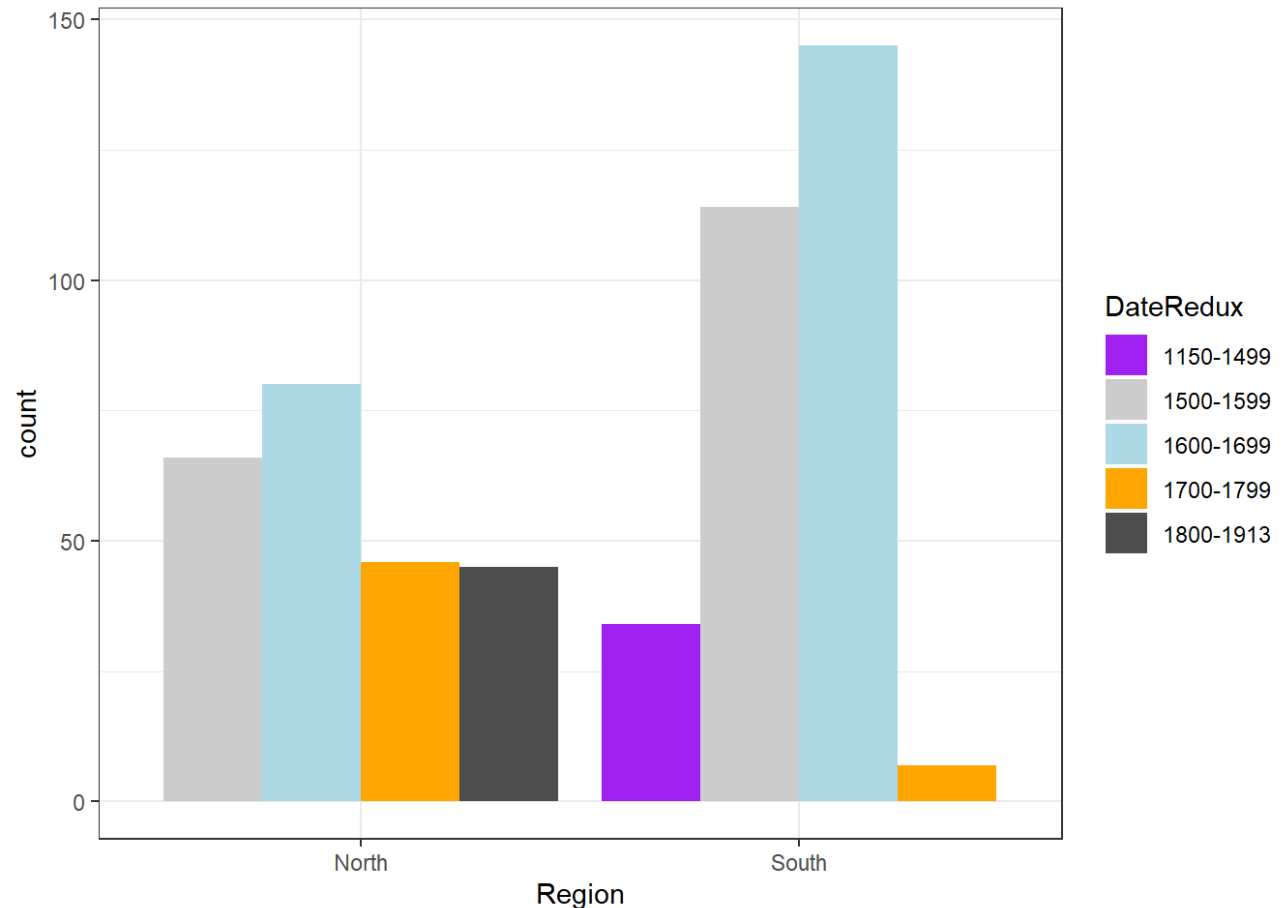

```
ggplot(pdat, aes(Region, fill = DateRedux)) +
    geom_bar(position = position_dodge(), stat = "count") +
    theme_bw() +
    scale_fill_manual(values = clrs)
```

To create **grouped** bar plots, we plot Region while including DateRedux as **the fill argument**.


```
ggplot(bdat, aes(DateRedux, Percent, fill = DateRedux)) +
    geom_bar(stat="identity") + theme_bw() +
        geom_text(aes(y = Percent-5, label = Percent), color = "white", size=3) +
        scale_fill_manual(values = clrs)
```

Boxplots, or Box-and-Whisker Plots, are exploratory graphics and they show the relationships between **categorical** and **numeric** variables.
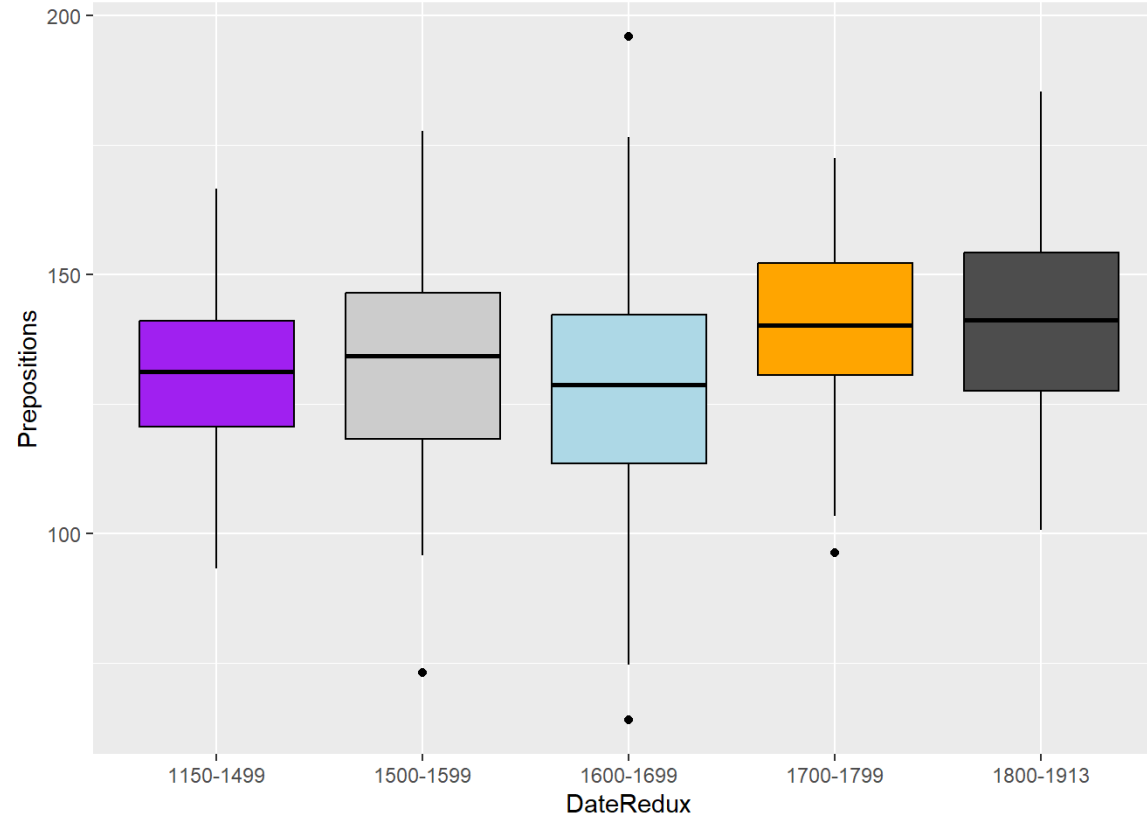

```
# create boxplot
ggplot(pdat, aes(DateRedux, Prepositions, color = GenreRedux)) +
geom_boxplot(fill=clrs, color="black")
```