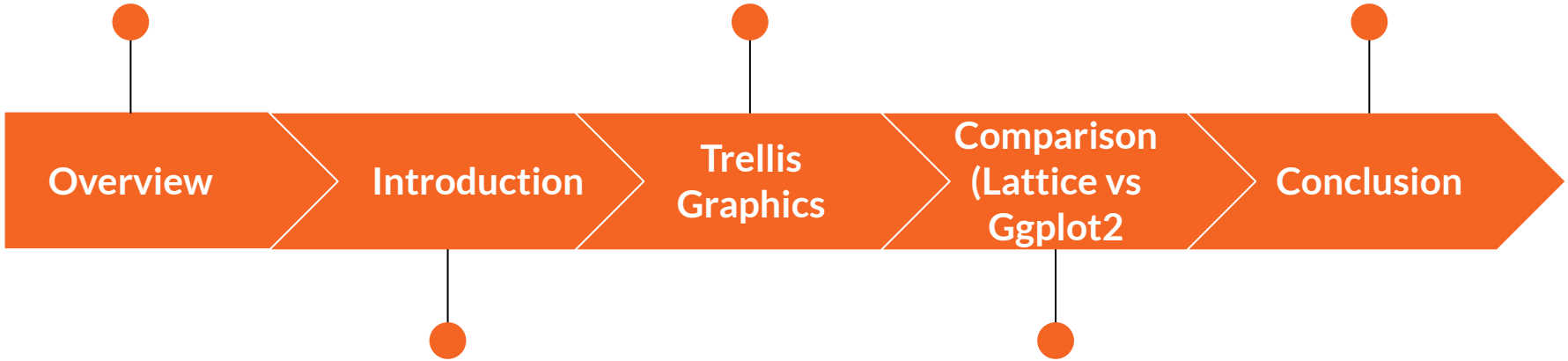
Two-Way-Dot Chart

MidTerm Teaching Presentation.

Isaac Kobby Anni • 10.22.2024 • Data Science, Ph.D

Presentation Contents

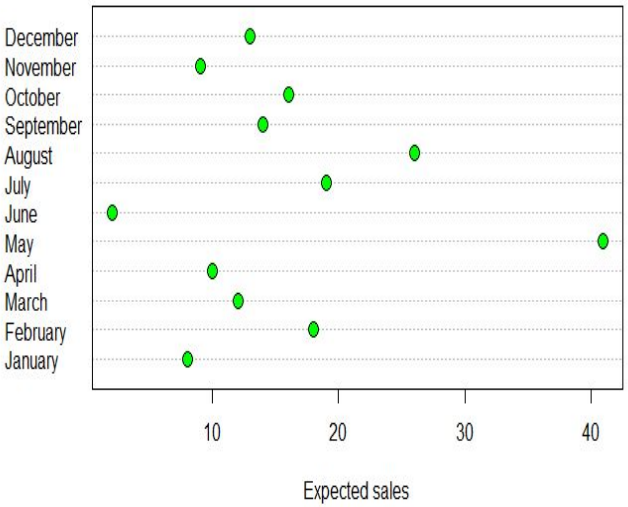


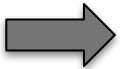
Overview

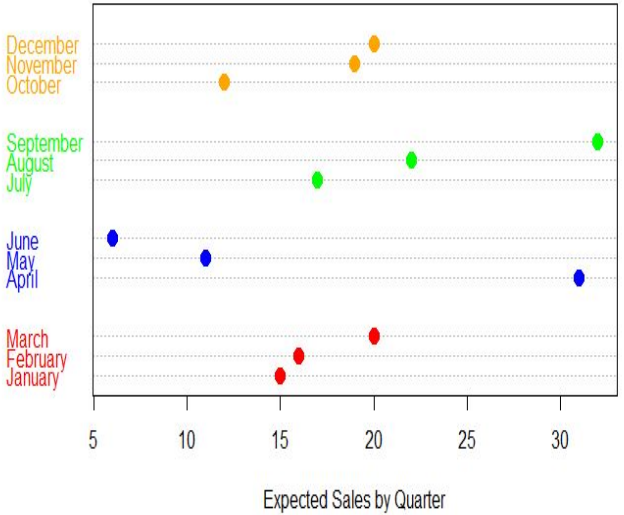
Chapter 8.5 - visualizing

Multivariable Data

Overview


December
November
October
September
August
July
June
May
April
March
February
January
10
20
30
40
Expected sales




December
November
October
September
August
July
June
May
April
March
February
January
5
10
15
20
25
30
Expected Sales by Quarter

Introduction to Two-Way-Dot Chart.

Key Takeaways

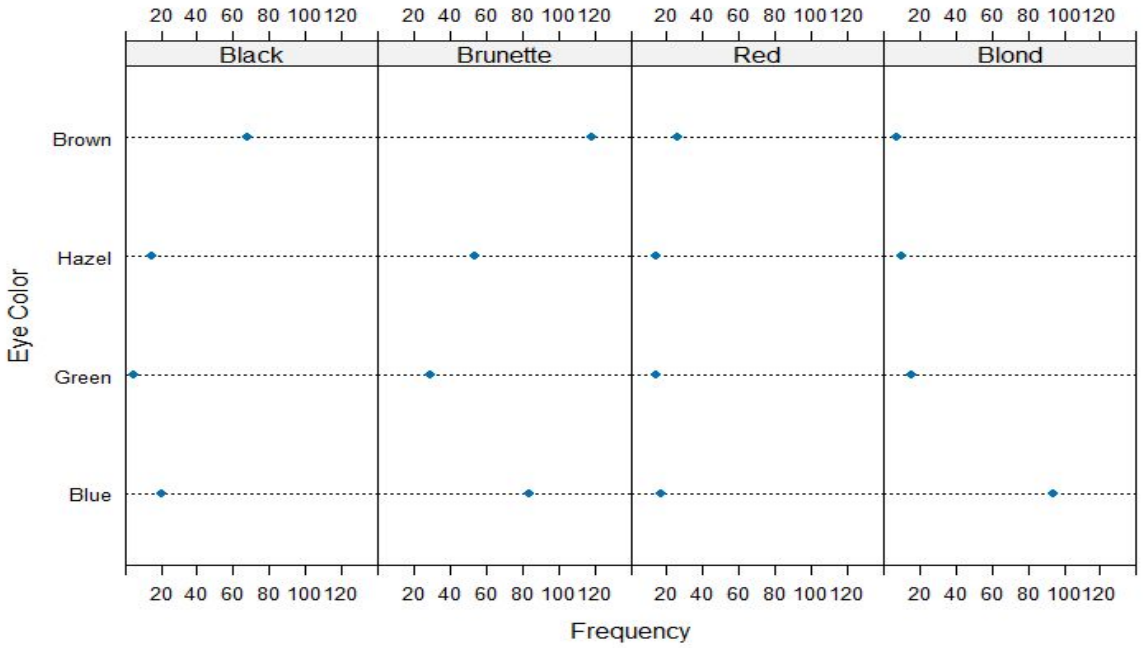
- Introduced by Cleveland (1985)
- Designed for comparing two categorical variables.
- Eg: Hair & Color data (Table 8.1)
- Each point represents a specific combination of the two variables.
- Used to visualize distributions and frequencies across two dimensions (2 variables).

Trellis Graphics Overview

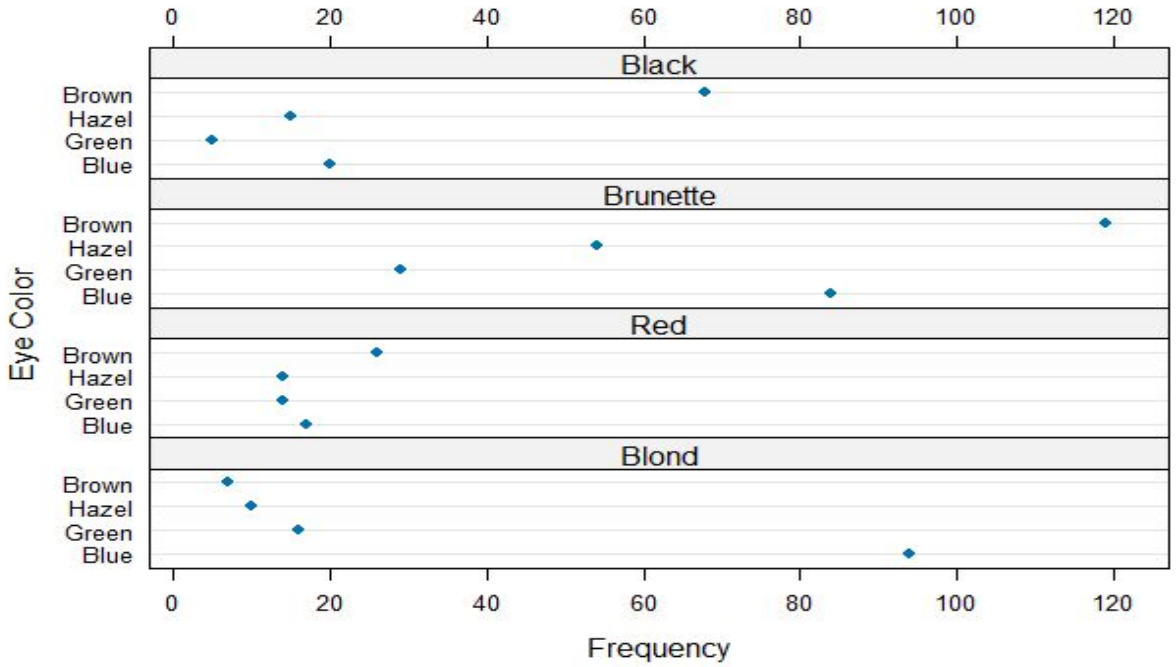
Facts.

- Trellis graphics is the implementation of multi-variable data in R using the lattice package.
- Developed to handle complex multi-variable data more easily.
- Used to visualize categorical data relationships.

Trellis Plot- Horizontal Layout



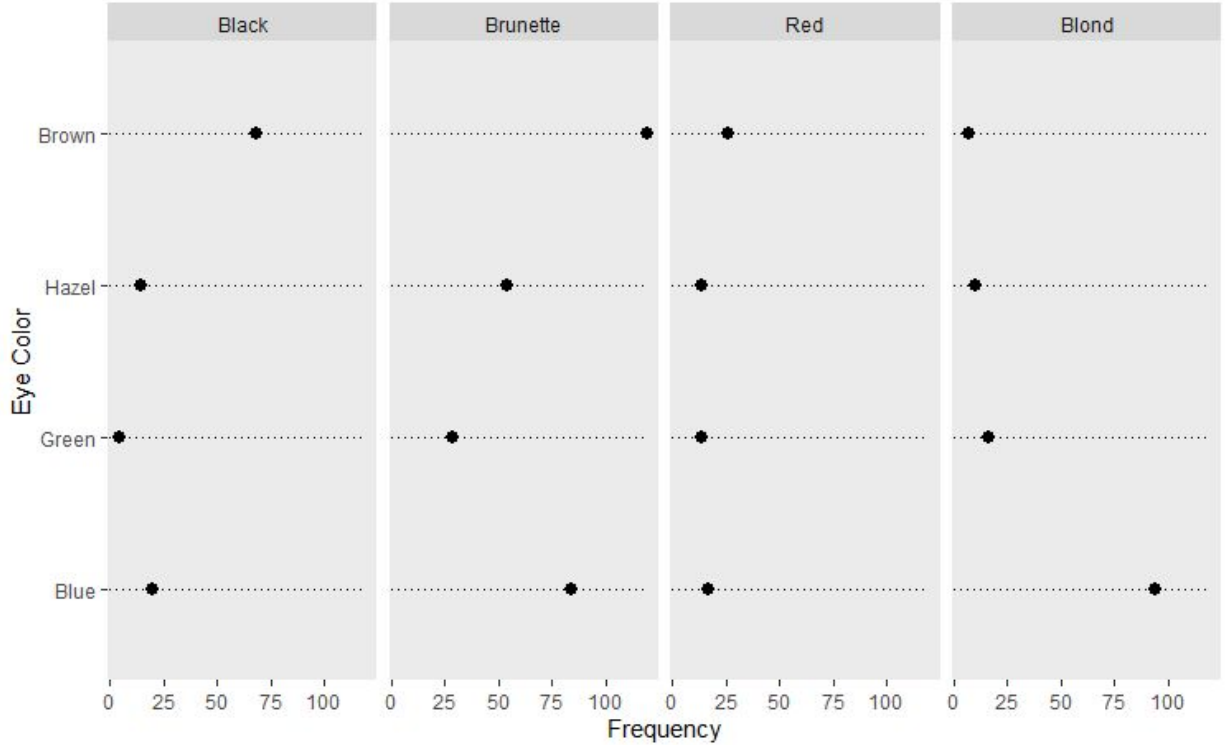
Trellis Plot- Vertical Layout



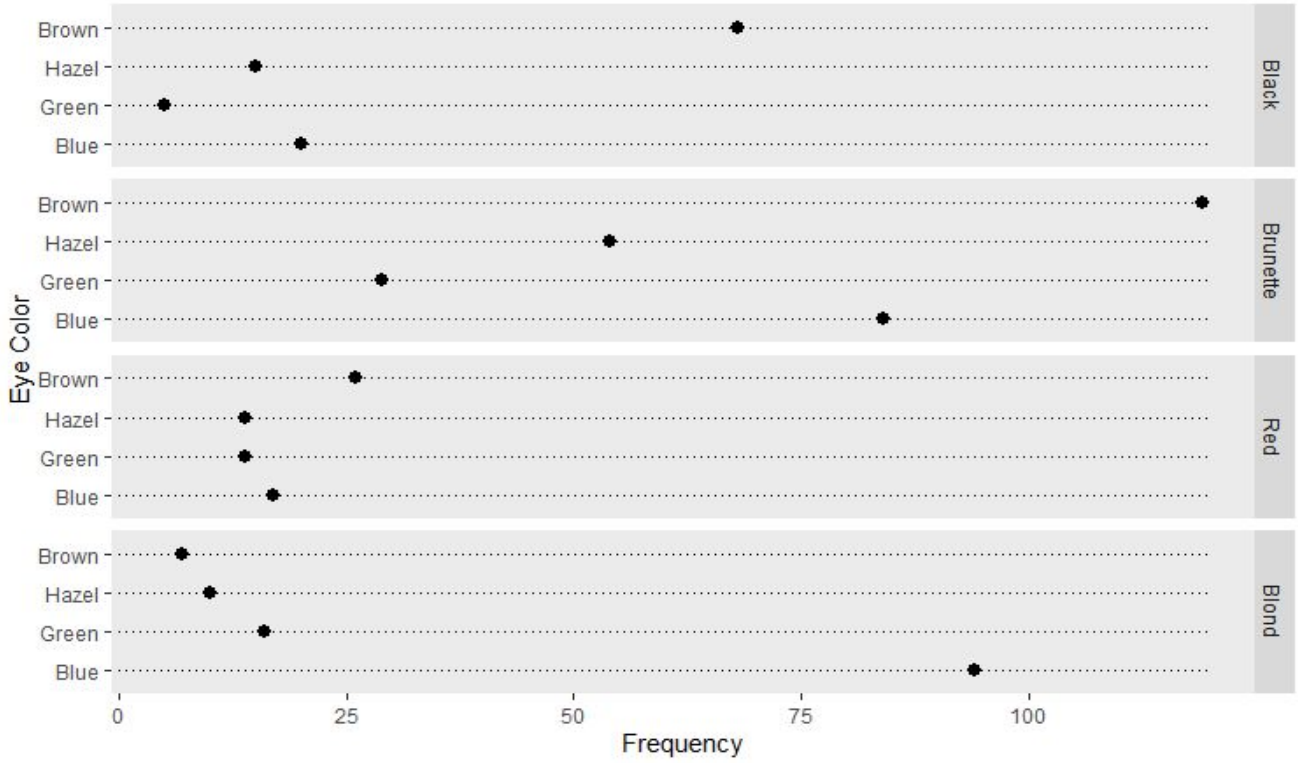
Trellis Plot - Code Snippet

```r
```{r}
haireye<-matrix(data=c(20,84,17,94,5,29,14,16,15,54,14,10,68,119,26,7),
                       nrow=4,
                       ncol=4,
                       byrow=TRUE,
                       dimnames=list(c("Blue","Green","Hazel","Brown"),
                       c("Black","Brunette","Red","Blond")))
figure<-dotplot(haireye,xlab="Frequency",
                    ylab="Eye Color",
                    as.table=TRUE,
                    groups=FALSE,
                    stack=FALSE,
                    layout=c(1,4),
                    scales=list(alternating=3))
print(figure)
```
```

Ggplot2 - Horizontal Layout



Ggplot2 - Vertical Layout



Ggplot2 - Code Snippet

```r
require(ggplot2)

graphics.off()
windows(width=4.5,height=2.0)

haireye<-matrix(data=c(7,10,16,94,26,14, 14,17,119,54,29,84,68,15,5,20),
                nrow=4,ncol=4,
                byrow=TRUE,
                dimnames=list(c("Blond", "Red","Brunette","Black"),
                         c("Brown","Hazel","Green","Blue"))
              )

haircol<-unlist(dimnames(haireye)[1])
eyecol<-unlist(dimnames(haireye)[2])

hair<-rep(" ",16)
eye<-rep(" ",16)
freq<-rep(0,16)
n=0
for (i in 1:4){ for (j in 1:4){
n<-n+1
hair[n]<-haircol[i]
eye[n]<-eyecol[j]
freq[n]<-haireye[i,j]
}}

hair<-factor(hair,levels=rev(haircol),ordered=TRUE)
eye<-factor(eye,levels=rev(eyecol),ordered=TRUE)

haireyef<-data.frame(hair=hair,eye=eye,freq=freq)

figure<-ggplot(haireyef,aes(x=eye,y=freq)) +
        geom_pointrange(ymin=0,ymax=120,linetype=3) +
        facet_grid( hair)+
        labs(x="Eye Color",y="Frequency") +
        theme(panel.grid=element_blank()) +
        coord_flip()

print(figure)
```

Comparison of Lattice & Ggplot2

Lattice

- Less flexible
- Used in Trellis
- Modification Restricted to Trellis usage

Ggplot2

- More flexible
- Standalone
- Depends on Ggplot2 flexibility

Challenges & Best practices

Challenges

- Horizontal Dot Chart compresses the x - axis.
- Ggplot2 requires more detailed coding to achieve desired layout

Best Practices

- R graphics package is simple and easy for two-way-dot chart
- Vertical layout is gives good layout for presentation.

Appendix - Flattened Rep.

```r
haireye <- matrix(data=c(20,84,17,94,5,29,14,16,15,54,
                14,10,68,119,26,7),
                nrow=4,
                ncol=4,
                byrow=TRUE,
                dimnames=list(c("Blue","Green","Hazel","Brown"),
                c("Black","Brunette","Red","Blond"))
                )
haireye
```

	Black	Brunette	Red	Blond
Blue	20	84	17	94
Green	5	29	14	16
Hazel	15	54	14	10
Brown	68	119	26	7

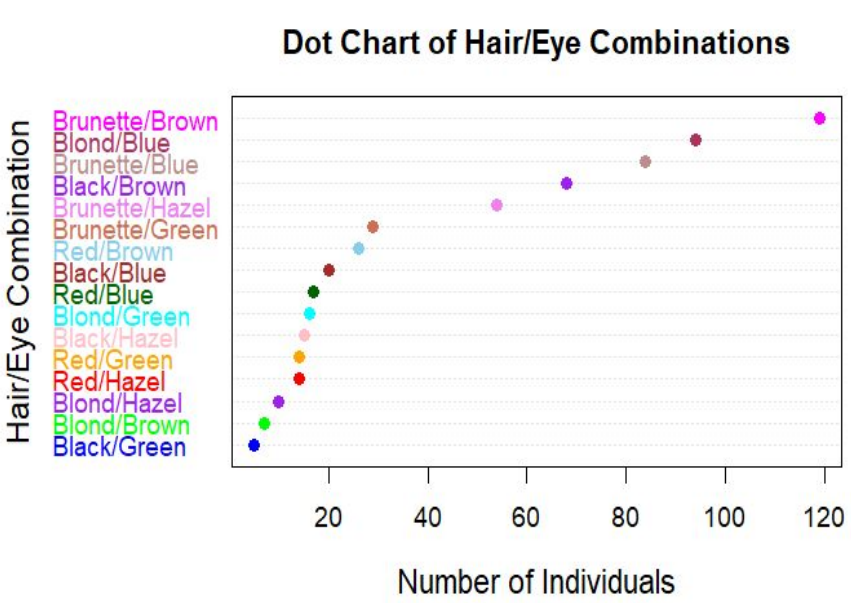
```r
flattened_data = as.data.frame(as.table(haireye))
flattened_data
```

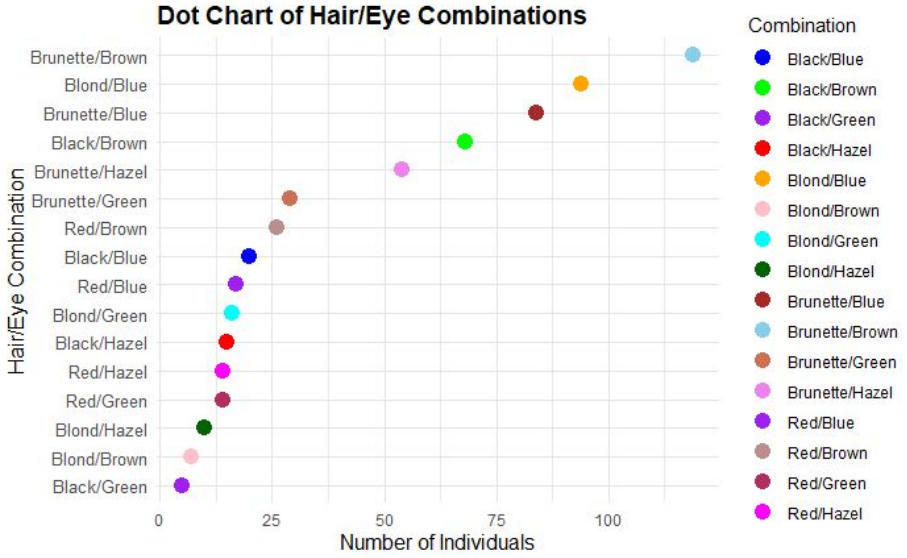
Description: df [16 × 3]

Var1 <fctr>	Var2 <fctr>	Freq <dbl>
Blond	Brown	7
Red	Brown	26
Brunette	Brown	119
Black	Brown	68
Blond	Hazel	10
Red	Hazel	14
Brunette	Hazel	54
Black	Hazel	15
Blond	Green	16
Red	Green	14

1-10 of 16 rows

R graphics vs Ggplot2 for flattened data





Questions?

Thank you for your attention.