

Atividade 03

Análise e Visualização de Dados

Victor Augusto Lima de Almeida
Recife, 15 de Setembro de 2020

1. Que outro tipo de gráfico poderíamos utilizar para avaliar dados categóricos?

R- De longe o método mais comum de apresentação de dados categóricos é o gráfico de barras, porém eu acredito que um gráfico que poderia ser utilizado para apresentar esse tipo de dados é o histograma assim como é feito com os dados numéricos, os histogramas se encaixam bem quando tentar demonstrar dados em classes ou categorias.

2. Como podemos representar de forma gráfica a distribuição de dados numéricos?

R- Acredito que a melhor forma de representar distribuições de dados gráficos é por meio de um gráfico de barras, pois, esse gráfico é consistente em comparar dados ou expor um desenvolvimento sendo ele crescente ou decrescente. (Excelente para mostrar uma boa ou má distribuição dos dados) Além disso, o gráfico de barras é as informações são distribuídas linearmente o que irá favorecer a interpretação da distribuição.

3. O que o método 'geom_smooth' da biblioteca ggplot2 faz?

R- A função `geom_smooth` facilita a inclusão de linhas de regressão. Se não for fornecido um argumento à função ela tentará uma aproximação **LOESS**.

4. Pesquisar e exemplificar o que é 'loess' e 'gam'

R- **LOESS** são regressões estimadas localmente, geralmente utilizadas quando há muito ruído, ou até quando a regressão

pelos mínimos quadrados se torna muito custosa. Utilizado na análise de dados como o de comportamento de eleitores.

- **GAMs** são modelos generalizados, no qual as variáveis preditivas são na verdade funções suaves não parametrizadas, ou seja, dependem apenas do conjunto de dados a ser trabalhado, e não de parâmetros passados. É aplicado em somas ponderadas de regressão linear, mas não é tão eficaz. Isso acontece porque o mundo real de forma geral não é linear, e os dados interagem entre si.