

# 新疆2019年风电发电

## 目录

|               |   |
|---------------|---|
| 导入数据          | 1 |
| 可视化           | 2 |
| 实际发电功率分布      | 2 |
| 按小时可视化发电功率    | 3 |
| 相关系数矩阵        | 4 |
| 风速的分布         | 5 |
| 风速与时间和发电功率的关系 | 9 |

## 导入数据

我们选取了国家电网集团新疆供电分公司的风电负荷数据集，包含包括湿度，风速，温度等多个变量和主发电负荷的35040个数据，具体时间跨度为十五分钟，也就是一天共96个数据点。随后，选取相应数据进行探索性分析和基础可视化，为后续工作打下基础。

```
library(readxl)
library(tidyverse)
library(knitr)
data <- read_excel("./data/新疆风电2019.xlsx")
# 仅选择前5列和前5行
data %>%
  select(1:5) %>% # 选择前5列
  head(5) %>%
  kable(caption = "新疆风电2019年数据")
```

表 1: 新疆风电2019年数据

| 时间                  | 实际发电功率<br>(mw) | 测风塔30m风<br>速(m/s) | 测风塔50m风<br>速(m/s) | 测风塔10m风<br>向(°) |
|---------------------|----------------|-------------------|-------------------|-----------------|
| 2019-01-01 00:00:00 | 0.979591       | 0                 | 0.000             | 166.816         |
| 2019-01-01 00:15:00 | 1.150984       | 0                 | 0.000             | 166.832         |
| 2019-01-01 00:30:00 | 1.066162       | 0                 | 0.000             | 166.859         |
| 2019-01-01 00:45:00 | 0.923717       | 0                 | 0.000             | 166.894         |
| 2019-01-01 01:00:00 | 0.813552       | 0                 | 0.297             | 166.892         |

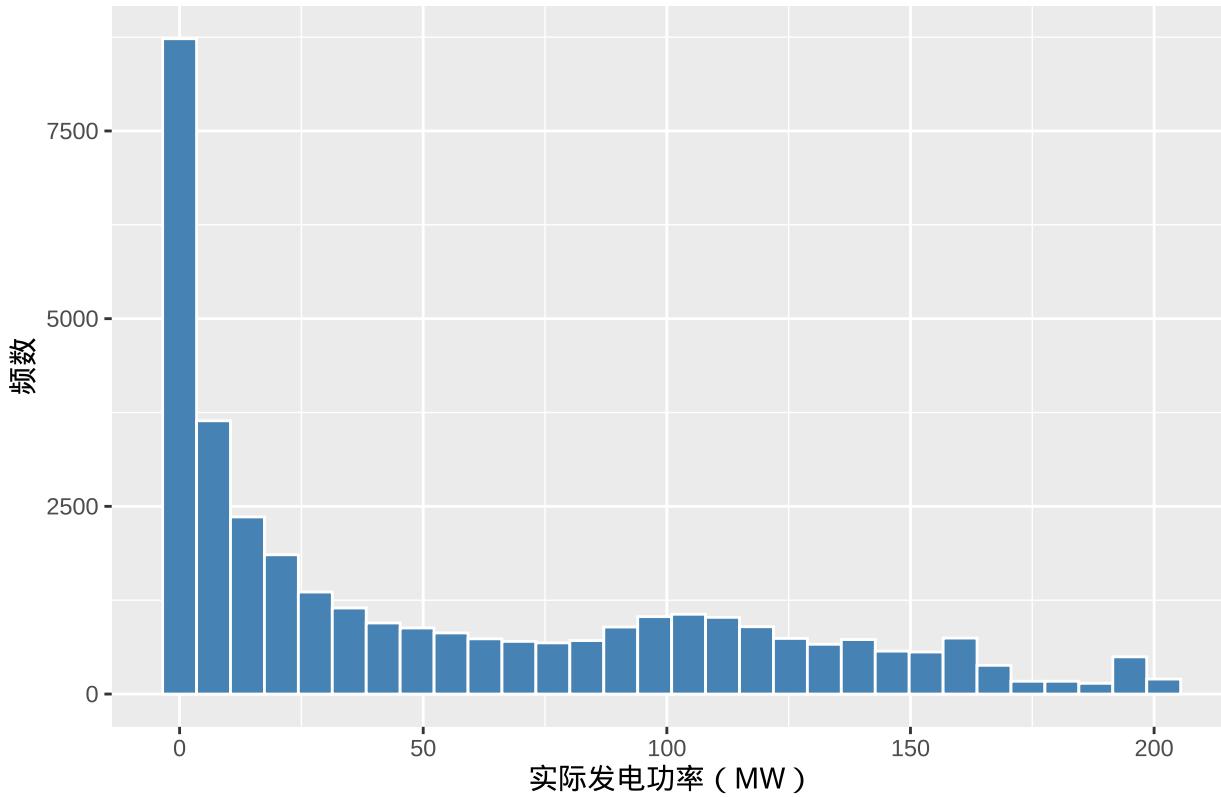
```
# 数据清理，使用mutate函数对dataframe类型创建新的列：时间，month, hour
data <- data %>%
  mutate(
    时间 = ymd_hms(时间),           # 转换时间为日期时间格式
    month = month(时间, label = TRUE), # 提取月份
    hour = hour(时间)               # 提取小时
  )
```

## 可视化

### 实际发电功率分布

```
# geom_histogram表示绘制直方图，bins为柱子的数量，fill表示填充颜色，color表示边框颜色
ggplot(data, aes(x = `实际发电功率 (mw)`)) +
  geom_histogram(bins = 30, fill = "steelblue", color = "white") +
  labs(
    title = "实际发电功率分布",
    x = "实际发电功率 (MW)",
    y = "频数"
  )
```

## 实际发电功率分布



该图展示了实际发电功率的分布情况，横轴为实际发电功率，纵轴为频数。从图中可以看出，发电功率主要集中在低功率范围，此区间的频数明显最高，特别是功率接近0时，频数达到峰值。

随着发电功率的增加，频数迅速下降，中高功率范围的出现频率较低，分布较为均匀。这种分布可能反映了实际运行中低功率状态的设备使用更为频繁，而高功率状态可能仅在需求峰值或特殊情况下出现。

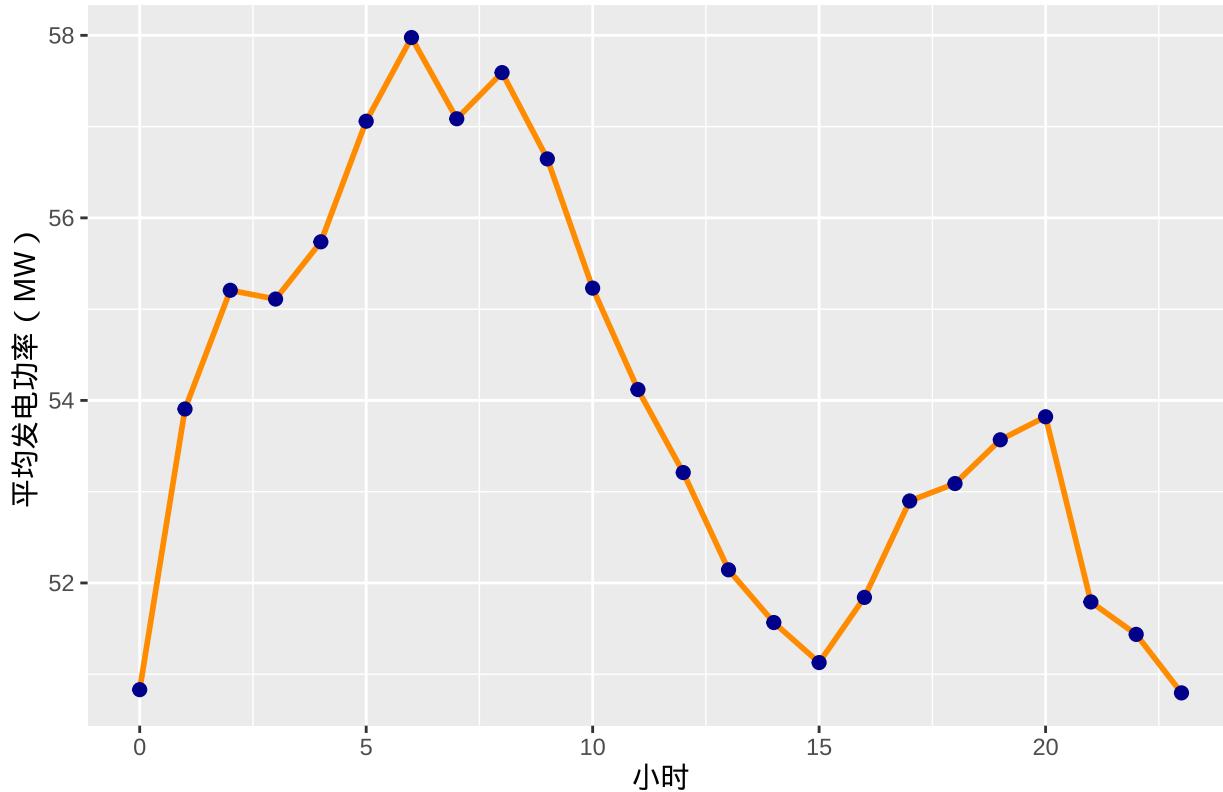
想研究环境对发电的影响，可以从时间、风速、湿度等方面研究

## 按小时可视化发电功率

```
# geom_line中设置线段的参数, geom_point中设置点的参数
hourly_data <- data %>%
  group_by(hour) %>%
  summarise(avg_power = mean(`实际发电功率 (mw)`), na.rm = TRUE)
ggplot(hourly_data, aes(x = hour, y = avg_power)) +
  geom_line(color = "darkorange", size = 1) +
  geom_point(color = "darkblue", size = 2) +
  labs(
    title = "一天中不同小时的平均发电功率",
    x = "小时",
```

```
y = "平均发电功率 (MW)"  
)
```

### 一天中不同小时的平均发电功率



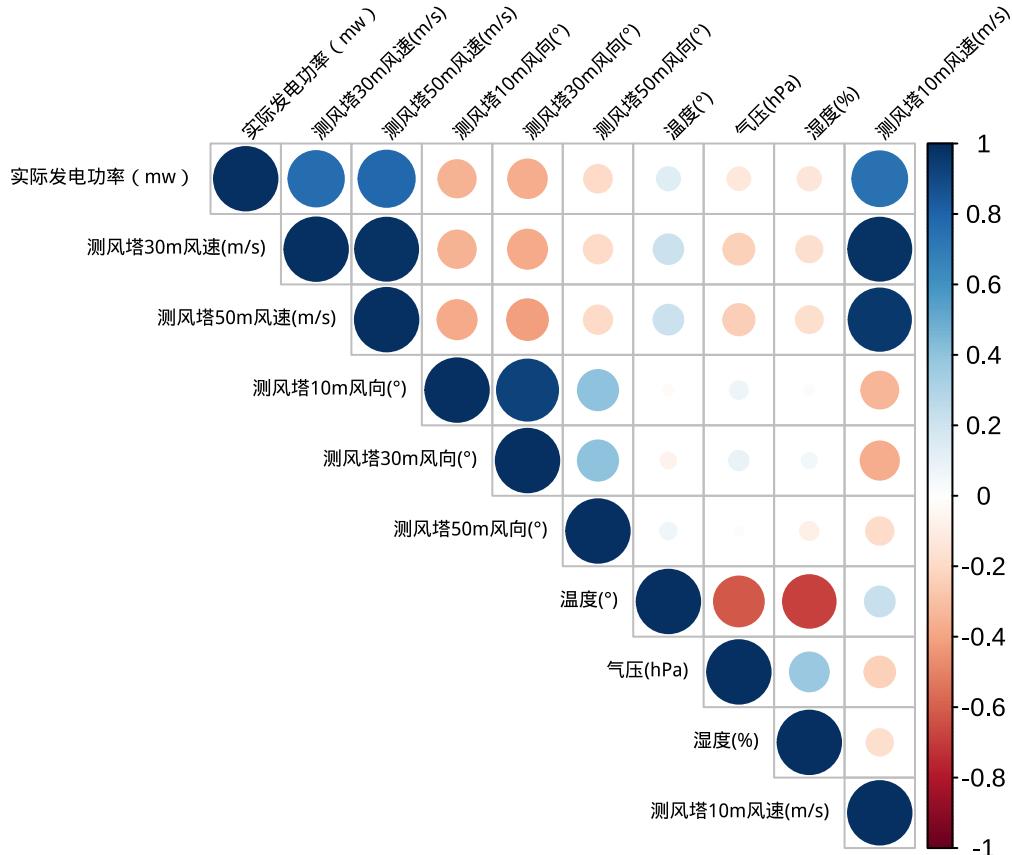
图中展示了一天24小时内发电功率的变化趋势，呈现出典型的日负荷模式。

凌晨至早晨（0-6时）功率逐渐上升，早晨达到高峰；上午至中午（6-12时）功率开始下降；下午继续下降，达到最低点；晚上时段（18-22时）功率回升，并在20时左右达到次高峰；深夜（22-24时）功率迅速下降，进入最低点。早晚高峰反映了居民和社会活动的规律，午后低谷和深夜低谷则与人们活动减少和设备关闭有关。

### 相关系数矩阵

```
library(reshape2)  
library(corrplot)  
# 计算相关系数矩阵  
data1 <- read_excel("./data/新疆风电2019.xlsx")  
# 选择数值型列计算相关系数矩阵  
numerical_columns <- data1[sapply(data1, is.numeric)]  
  
# 计算相关系数矩阵
```

```
cor_matrix <- cor(numerical_columns)
# 使用 corrplot 包绘制相关系数矩阵
corrplot(cor_matrix, method = "circle", type = "upper", tl.col = "black", tl.srt = 45, tl.cex=0.6)
```

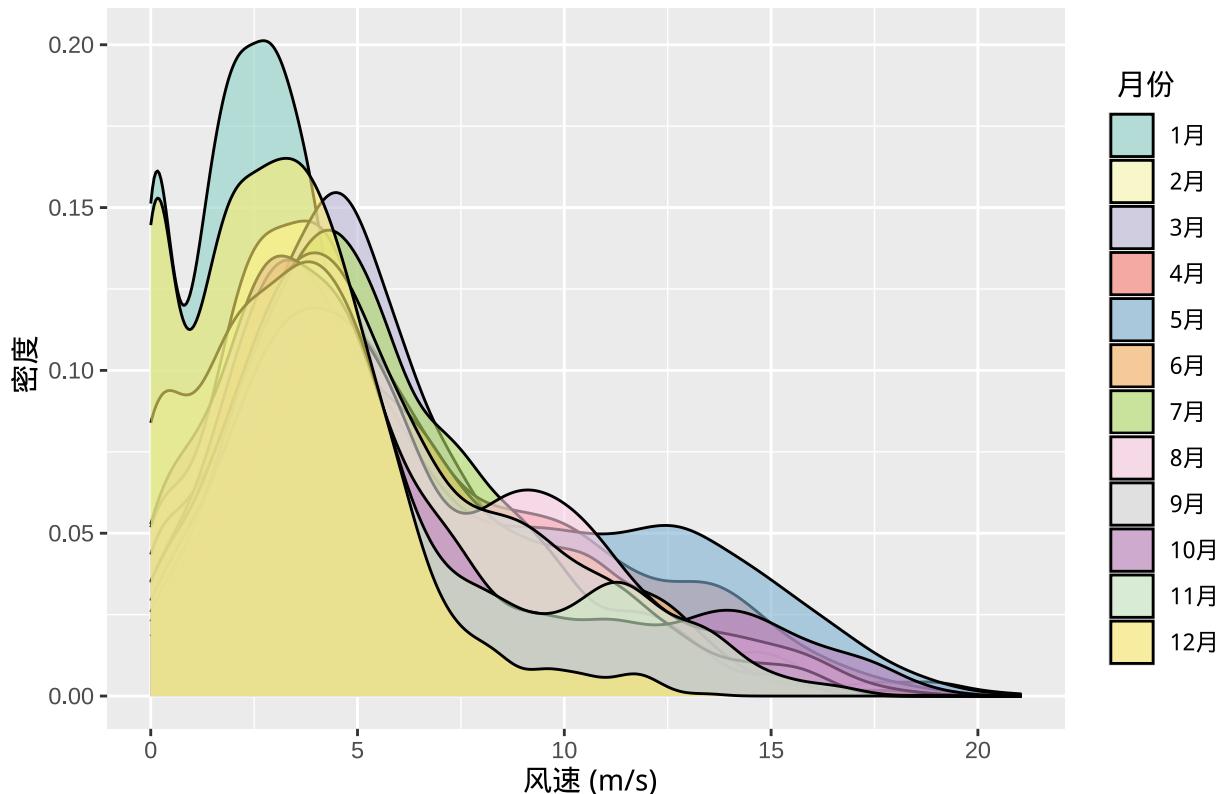


发电功率和风速关系较大，而和其他属性相关性很小，我们可以着重分析发电功率和风速。

## 风速的分布

```
# geom_density 表示绘制密度图, alpha 表示透明度
# scale_fill_brewer 表示用合适的颜色为每个月份的密度图绘制颜色
ggplot(data, aes(x = `测风塔30m风速(m/s)`, fill = month)) +
  geom_density(alpha = 0.6) +
  scale_fill_brewer(palette = "Set3") +
  labs(
    title = "不同月份 30 米风速分布",
    x = "风速 (m/s)",
    y = "密度",
    fill = "月份"
  )
```

## 不同月份 30 米风速分布



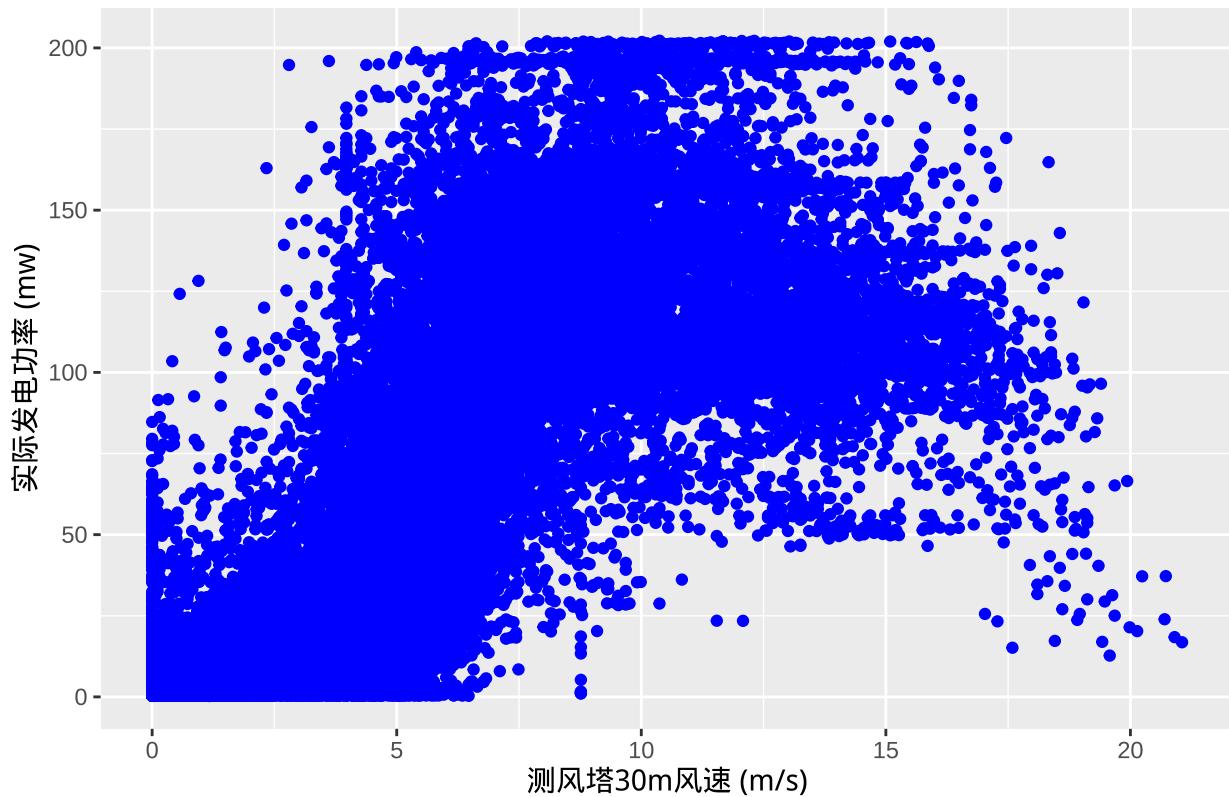
图中展示了不同月份的30米高度风速分布特征，风速大多集中在 $0\sim 5 \text{ m/s}$ 区间，表明低风速发生频率较高，而高风速( $>10 \text{ m/s}$ )则较为罕见。

冬季风速分布较窄，集中在低风速区间，反映了天气系统稳定；夏季风速分布较宽，较高风速的概率增大，可能与夏季的对流活动增强有关；春秋季节风速分布介于两者之间，表现出过渡特征。

随着季节变化，风速分布逐渐扩展，尤其是高风速的发生概率在春夏季略有增加。

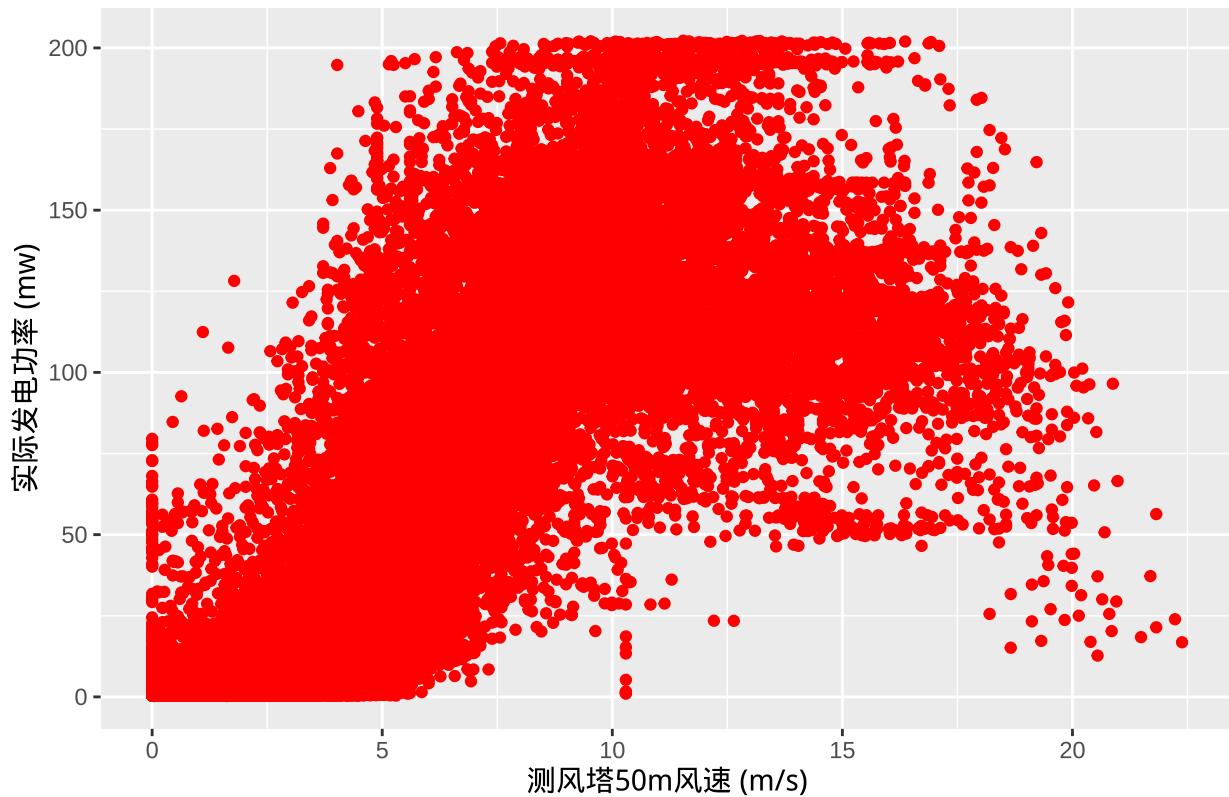
```
library(reshape2)
# 选择与发电功率和风速相关的列
data_clean <- data[, c("实际发电功率 (mw)", "测风塔30m风速(m/s)", "测风塔50m风速(m/s)", "测风塔10m风速(m/s)")]
cor_matrix <- cor(data_clean)
# 绘制不同风速与发电功率之间的关系图
# 1. 测风塔30m风速与实际发电功率的关系
ggplot(data, aes(x = `测风塔30m风速(m/s)`, y = `实际发电功率 (mw)`)) +
  geom_point(color = "blue") +
  labs(title = "测风塔30m风速与实际发电功率的关系",
       x = "测风塔30m风速 (m/s)",
       y = "实际发电功率 (mw)")
```

### 测风塔30m风速与实际发电功率的关系



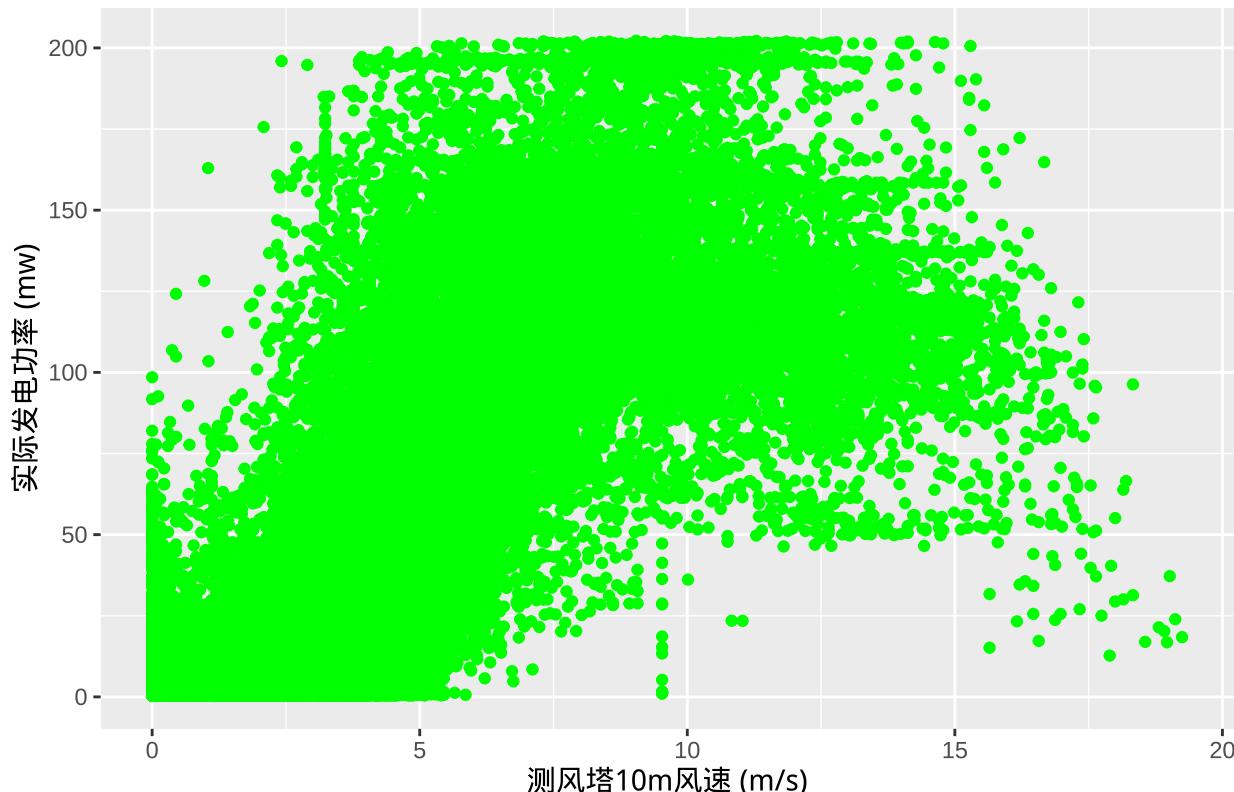
```
# 2. 测风塔50m风速与实际发电功率的关系
ggplot(data, aes(x = `测风塔50m风速(m/s)` , y = `实际发电功率 (mw)`)) +
  geom_point(color = "red") +
  labs(title = "测风塔50m风速与实际发电功率的关系",
       x = "测风塔50m风速 (m/s)",
       y = "实际发电功率 (mw)")
```

### 测风塔50m风速与实际发电功率的关系



```
# 3. 测风塔10m风速与实际发电功率的关系
ggplot(data, aes(x = `测风塔10m风速(m/s)`, y = `实际发电功率 (mw)`)) +
  geom_point(color = "green") +
  labs(title = "测风塔10m风速与实际发电功率的关系",
       x = "测风塔10m风速 (m/s)",
       y = "实际发电功率 (mw)")
```

## 测风塔10m风速与实际发电功率的关系



这三

幅图是不同高度的测风塔，风速与发电功率的关系。

我们发现并不是风速越快发电功率越高，可能是因为过快的风速会对设备造成损害，或者发电过多造成浪费。

最佳区间在5-10m/s左右。

## 风速与时间和发电功率的关系

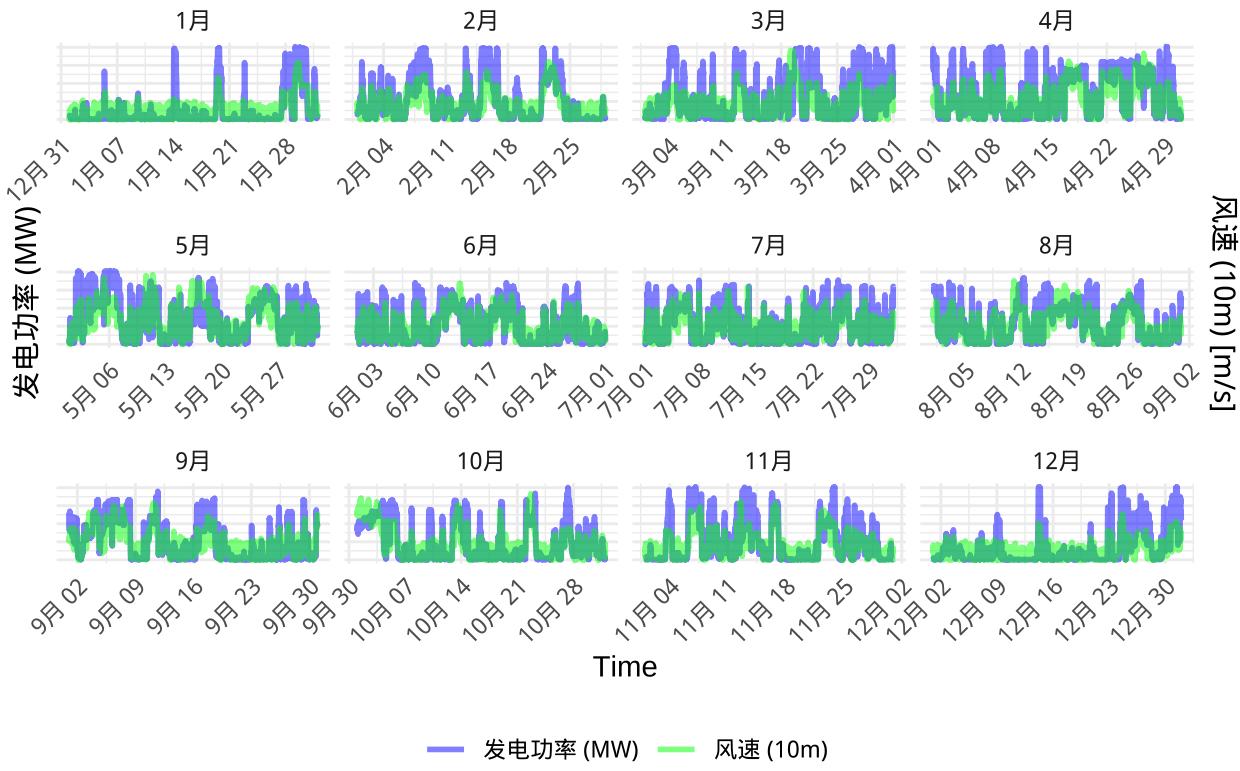
```
# 绘制双Y轴图
ggplot(data, aes(x = 时间)) +
  # 绘制发电功率曲线，红色
  geom_line(aes(y = `实际发电功率 (mw)`), color = "发电功率 (MW)", size = 1, alpha = 0.5) +
  # 绘制风速曲线，绿色，风速数值乘以10进行缩放，以便与发电功率在同一y轴范围内显示
  geom_line(aes(y = `测风塔10m风速(m/s)` * 10, color = "风速 (10m)"), size = 1, alpha = 0.5) +
  # 设置左侧y轴（发电功率）和右侧y轴（风速）
  scale_y_continuous(
    name = "发电功率 (MW)",
    sec.axis = sec_axis(
      trans = ~ . / 10, # 将右侧y轴数值恢复到原始风速值
    )
  )
```

```

    name = "风速 (10m) [m/s]"
)
) +
facet_wrap(~ month, scales = "free_x") +
labs(x = "Time", title = "每月的风速和发电功率关系") +
scale_color_manual(values = c("发电功率 (MW)" = "blue", "风速 (10m)" = "green")) +
theme_minimal() +
theme(
  axis.text.x = element_text(angle = 45, hjust = 1),
  axis.text.y = element_blank(), # 隐藏左侧Y轴的数值
  axis.text.y.right = element_blank(), # 隐藏右侧Y轴的数值
  legend.title = element_blank(),
  legend.position = "bottom", # 将图例放在下面
  legend.box = "horizontal" # 图例的布局为横向
)

```

## 每月的风速和发电功率关系



从图中可以看出，风速和发电功率之间存在明显的正相关关系，通常风速较高时，发电功率也较大。

然而，风速与发电功率之间的关系并非完全线性，尤其是在风速变化的极端范围内，发电功率的变化并不会完全按照风速的增加而线性增长，这可能与风力发电系统的功率曲线和效率限制有关。

图中的季节性变化也很明显，某些月份风速较大，导致发电功率随之增加，而其他月份风速较低时，发电功率相应减少。此外，图中还反映了风速和发电功率的波动性，尤其在某些时间点，风速的剧烈波动直接影响了发电功率的波动，这可能由气候变化或设备因素引起。

总体来看，风速和发电功率之间的关系可以帮助优化风力发电系统的效率，同时也揭示了风速变化对系统稳定性的重要影响，尤其是在风速较低或极端变化的情况下，可能需要采取额外的调节措施。