

데이터 분석 & 시각화

1. 데이터 분석 주제

주제 : 평균 기온, 강수량에 의한 교통량 변화 측정

주제 선정 이유

- 교통량은 다양한 변수들에 의해 크게 바뀐다. 일반적으로 생각하면 눈이 오는 날은 교통량이 적을 것이고, 명절이나 공휴일은 교통량이 증가할 것이다. 그렇다면 과연 기온과 강수량은 교통량에 유의미한 변화를 주는 지를 알아본다.

2. 데이터 파일 정보

분석 데이터 정보

1. 수원시 도로 교통량

- 2019/01/01 ~ 2019/12/31 기간의 경수대로, 덕영대로, 중부대로 상하행선 교통량 측정 데이터
- 데이터 수 : 2142개
- 출처 : 공공데이터포털(<https://www.data.go.kr/data/15080202/fileData.do>)
- csv 파일 구성요소

수집일자	교통량이 수집 된 날짜. 2019/01/01~2019/12/31 로 구성
수집지점	경수대로, 덕영대로, 중부대로
도로방향	상/하행선 구분
교통량	하루동안의 수집지점의 교통량의 수

수집일자	수집지점	도로방향	교통량(대)	2019-01-01 덕영대로(수원역열)	서울~수원	19873	2019-01-01 중부대로(신갈고가교~영덕고가차로 사이)	수원~용인	25221
2019-01-01	경수대로(시정앞 사거리~인계사거리)	수원~서울	19476	2019-01-02 덕영대로(수원역열)	서울~수원	24048	2019-01-02 중부대로(신갈고가교~영덕고가차로 사이)	수원~용인	38925
2019-01-02	경수대로(시정앞 사거리~인계사거리)	수원~서울	24663	2019-01-03 덕영대로(수원역열)	서울~수원	24122	2019-01-03 중부대로(신갈고가교~영덕고가차로 사이)	수원~용인	40876
2019-01-03	경수대로(시정앞 사거리~인계사거리)	수원~서울	25449	2019-01-04 덕영대로(수원역열)	서울~수원	24713	2019-01-04 중부대로(신갈고가교~영덕고가차로 사이)	수원~용인	40994
2019-01-04	경수대로(시정앞 사거리~인계사거리)	수원~서울	24955	2019-01-05 덕영대로(수원역열)	서울~수원	24417	2019-01-05 중부대로(신갈고가교~영덕고가차로 사이)	수원~용인	37561
2019-01-05	경수대로(시정앞 사거리~인계사거리)	수원~서울	24439	2019-01-06 덕영대로(수원역열)	서울~수원	20161	2019-01-06 중부대로(신갈고가교~영덕고가차로 사이)	수원~용인	28440
2019-01-06	경수대로(시정앞 사거리~인계사거리)	수원~서울	21134	2019-01-07 덕영대로(수원역열)	서울~수원	24383	2019-01-07 중부대로(신갈고가교~영덕고가차로 사이)	수원~용인	41570
2019-01-07	경수대로(시정앞 사거리~인계사거리)	수원~서울	25002	2019-01-08 덕영대로(수원역열)	서울~수원	24402	2019-01-08 중부대로(신갈고가교~영덕고가차로 사이)	수원~용인	42693
2019-01-08	경수대로(시정앞 사거리~인계사거리)	수원~서울	25474	2019-01-09 덕영대로(수원역열)	서울~수원	24577	2019-01-09 중부대로(신갈고가교~영덕고가차로 사이)	수원~용인	42559
2019-01-09	경수대로(시정앞 사거리~인계사거리)	수원~서울	25215	2019-01-10 덕영대로(수원역열)	서울~수원	24810	2019-01-10 중부대로(신갈고가교~영덕고가차로 사이)	수원~용인	42128
2019-01-10	경수대로(시정앞 사거리~인계사거리)	수원~서울	25424						

2. 수원시 평균 온도

- 수원시 2019년도 평균 기온 데이터
- 데이터 수 : 12개
- 출처 : 기상청 기상자료개방포털(<https://data.kma.go.kr/data/grnd/selectAsosRltmList.do?pgmNo=36>)
- csv 파일 구성요소

지점명	수원
일시	2019년 1월~12월
평균기온(°C)	월별 기온 평균값
평균최고기온(°C)	월별 최고기온 평균값
평균최저기온(°C)	월별 최저기온 평균값

지점명	일시	평균기온(°C)	평균최고기온(°C)	평균최저기온(°C)
수원	Jan-19	-1.1	4.9	-6.3
수원	Feb-19	0.8	6.8	-4.3
수원	Mar-19	6.6	13.3	0.9
수원	Apr-19	11.7	18.2	5.7
수원	May-19	18.4	25.9	11.8
수원	Jun-19	22.1	27.9	17.2
수원	Jul-19	25.8	30.3	22.3
수원	Aug-19	26.8	31.5	23.1
수원	Sep-19	22.3	26.8	18.3
수원	Oct-19	16	21.3	10.9
수원	Nov-19	7.5	13.4	2.2
수원	Dec-19	1.5	5.8	-3

3. 수원시 평균 강수량

- 수원시 2019년도 평균 강수량 데이터
- 데이터 수 : 12개
- 출처 : 기상청 기상자료개방포털(<https://data.kma.go.kr/data/grnd/selectAsosRltmList.do?pgmNo=36>)
- csv 파일 구성요소

지점명	수원
일시	2019년 1월~12월
월합강수량(mm)	월별 강수량 합
일최다강수량(mm)	월별 하루 최다강수량

지점명	일시	월합강수량(00~24h만)(mm)	일최다강수량(mm)
수원	Jan-19	0.5	0.4
수원	Feb-19	33.4	22.7
수원	Mar-19	39.6	23.6
수원	Apr-19	43.6	18.1
수원	May-19	26.7	16.5
수원	Jun-19	68.8	38.7
수원	Jul-19	190.9	108
수원	Aug-19	117.9	46.7
수원	Sep-19	201.9	80.6
수원	Oct-19	73.1	25.6
수원	Nov-19	93.2	41.6
수원	Dec-19	26.2	6.2

3. 결과 데이터

- 2019년 월별 교통량, 평균온도, 월강수량 데이터로 추출
- 데이터 수 : 36개

```
<SQL 코드>
SET HEADING ON -- 헤더 ON
SET PAGESIZE 100 -- 한페이지에 표시할 레코드 수
SET TERM OFF -- 조회결과 표시 OFF
SET COLSEP ' ' -- 데이터 구분자
SET FEEDBACK OFF -- 결과의 데이터건수, 시간표시 OFF

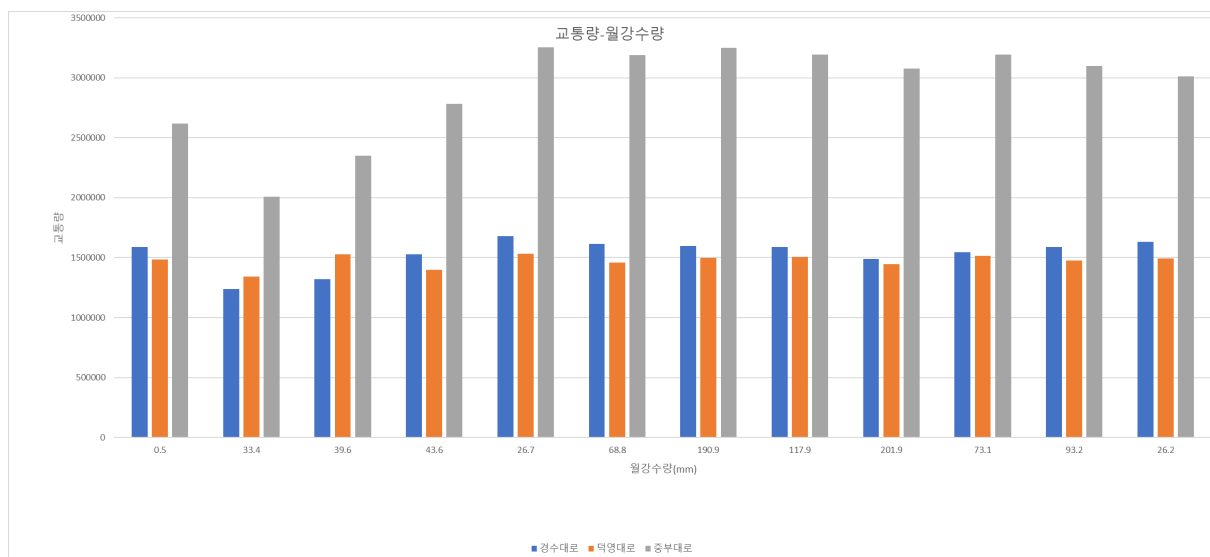
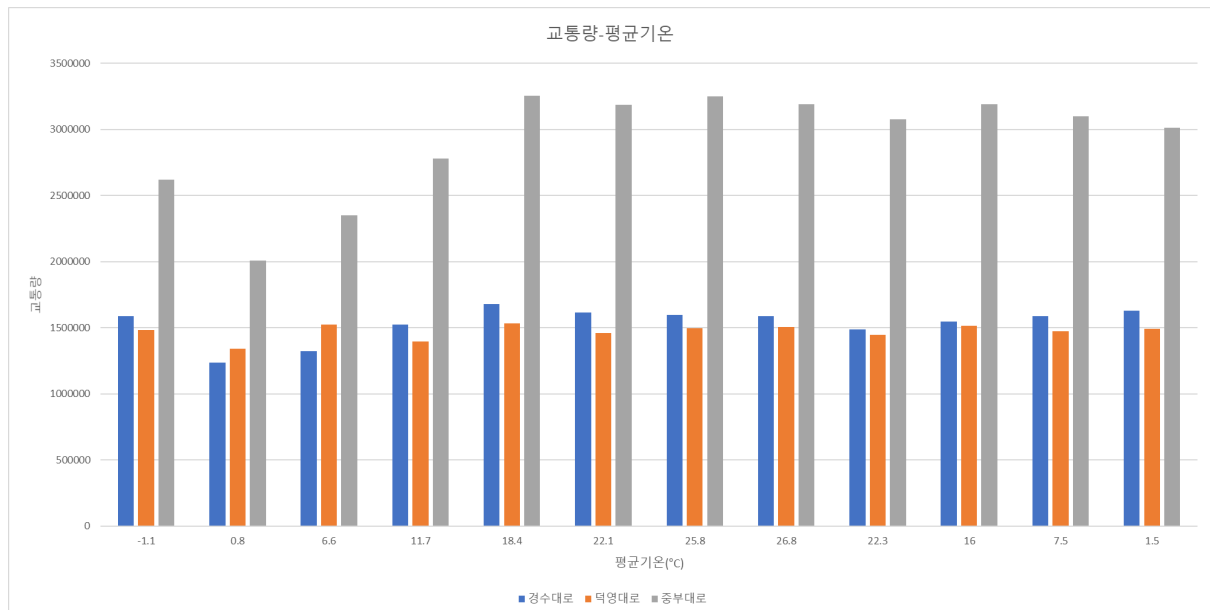
SPOOL C:\SQLDEV\TEST1.CSV -- 저장할 CSV 파일 이름, 위치 설정
SELECT TO_CHAR(T.MDATE, 'YYYY/MM/DD') AS "년/월"
      , T.REGION AS 지역
      , TRA.지점
      , TRA.교통량
      , T.AVGTEMPERATURE AS 평균온도
      , P.MONPRECIPITATION AS 월강수량
FROM TEMPERATURE T
INNER JOIN PRECIPITATION P
ON T.MDATE = P.MDATE
LEFT OUTER JOIN (SELECT TO_CHAR(COLDATE, 'YY/MM') AS 년월
                  , COLPOINT AS 지점
                  , SUM(VOLTRAFFIC) AS 교통량
FROM TRAFFIC
GROUP BY TO_CHAR(COLDATE, 'YY/MM'), COLPOINT
ORDER BY TO_CHAR(COLDATE, 'YY/MM') ASC, COLPOINT DESC) TRA
ON TO_CHAR(P.MDATE, 'YY/MM') = TRA.년월
ORDER BY TRA.지점 ASC;
SPOOL OFF
```

년/월	지역	지점	교통량	평균온도	월강수량
1 2019/01/01	수원	경수대로 (시청앞 사거리-민계사거리)	1589152	-1.1	0.5
2 2019/02/01	수원	경수대로 (시청앞 사거리-민계사거리)	1239087	0.8	33.4
3 2019/03/01	수원	경수대로 (시청앞 사거리-민계사거리)	1322618	6.6	39.6
4 2019/04/01	수원	경수대로 (시청앞 사거리-민계사거리)	1526837	11.7	43.6
5 2019/05/01	수원	경수대로 (시청앞 사거리-민계사거리)	1681014	18.4	26.7
6 2019/06/01	수원	경수대로 (시청앞 사거리-민계사거리)	1616647	22.1	68.8
7 2019/07/01	수원	경수대로 (시청앞 사거리-민계사거리)	1595819	25.8	190.9
8 2019/08/01	수원	경수대로 (시청앞 사거리-민계사거리)	1589671	26.8	117.9
9 2019/09/01	수원	경수대로 (시청앞 사거리-민계사거리)	1489024	22.3	201.9
10 2019/10/01	수원	경수대로 (시청앞 사거리-민계사거리)	1546779	16	73.1
11 2019/11/01	수원	경수대로 (시청앞 사거리-민계사거리)	1589753	7.5	93.2
12 2019/12/01	수원	경수대로 (시청앞 사거리-민계사거리)	1629933	1.5	26.2
13 2019/01/01	수원	덕영대로 (수원역앞)	1484349	-1.1	0.5
14 2019/02/01	수원	덕영대로 (수원역앞)	1342642	0.8	33.4
15 2019/03/01	수원	덕영대로 (수원역앞)	1526180	6.6	39.6
16 2019/04/01	수원	덕영대로 (수원역앞)	1396535	11.7	43.6
17 2019/05/01	수원	덕영대로 (수원역앞)	1534025	18.4	26.7
18 2019/06/01	수원	덕영대로 (수원역앞)	1459904	22.1	68.8
19 2019/07/01	수원	덕영대로 (수원역앞)	1498133	25.8	190.9
20 2019/08/01	수원	덕영대로 (수원역앞)	1505413	26.8	117.9
21 2019/09/01	수원	덕영대로 (수원역앞)	1446632	22.3	201.9
22 2019/10/01	수원	덕영대로 (수원역앞)	1516818	16	73.1
23 2019/11/01	수원	덕영대로 (수원역앞)	1476310	7.5	93.2
24 2019/12/01	수원	덕영대로 (수원역앞)	1494322	1.5	26.2
25 2019/01/01	수원	중부대로 (신갈고가교-영덕고가차로 사이)	2619340	-1.1	0.5
26 2019/02/01	수원	중부대로 (신갈고가교-영덕고가차로 사이)	2007051	0.8	33.4
27 2019/03/01	수원	중부대로 (신갈고가교-영덕고가차로 사이)	2349229	6.6	39.6
28 2019/04/01	수원	중부대로 (신갈고가교-영덕고가차로 사이)	2781905	11.7	43.6
29 2019/05/01	수원	중부대로 (신갈고가교-영덕고가차로 사이)	3254252	18.4	26.7
30 2019/06/01	수원	중부대로 (신갈고가교-영덕고가차로 사이)	3188622	22.1	68.8
31 2019/07/01	수원	중부대로 (신갈고가교-영덕고가차로 사이)	3249004	25.8	190.9
32 2019/08/01	수원	중부대로 (신갈고가교-영덕고가차로 사이)	3192880	26.8	117.9
33 2019/09/01	수원	중부대로 (신갈고가교-영덕고가차로 사이)	3076898	22.3	201.9
34 2019/10/01	수원	중부대로 (신갈고가교-영덕고가차로 사이)	3191933	16	73.1
35 2019/11/01	수원	중부대로 (신갈고가교-영덕고가차로 사이)	3097869	7.5	93.2
36 2019/12/01	수원	중부대로 (신갈고가교-영덕고가차로 사이)	3012509	1.5	26.2

4. 데이터 시각화 & 결론

데이터 시각화

- Excel를 이용한 시각화



- Java를 이용한 시각화

```
<%@ page language="java" contentType="text/html; charset=UTF-8"
    pageEncoding="UTF-8"%>
<!DOCTYPE html>
<html>
<head>
<meta charset="UTF-8">
<title>Insert title here</title>
    <!--Load the AJAX API-->
    <script type="text/javascript" src="https://www.gstatic.com/charts/loader.js"></script>
</head>
<body>
    <%@ page import="java.sql.*"%>
    <h2>Column Chart Example</h2>

    <!--Div that will hold the pie chart-->
    <div id="columnchart_values"></div>

    <script type="text/javascript">
    google.charts.load("current", {packages:['corechart']});
    google.charts.setOnLoadCallback(drawChart);

    function drawChart() {

        var arr = new Array();
        arr.push(['Element', 'Density', { role: 'style' }]);

    <%
    try{
        String driverName = "oracle.jdbc.driver.OracleDriver";
        String dbURL="jdbc:oracle:thin:@192.168.119.119:1521:dink";

        Class.forName(driverName);
        Connection conn = DriverManager.getConnection(dbURL,"scott","tiger");

        Statement stmt = conn.createStatement();
        String sql = "SELECT TO_CHAR(T.MDATE, 'YYYY/MM/DD') AS 년월, T.REGION AS 지역, TRA.지점, TRA.교통량, T.AVGTEMPERATURE AS 평균온
        ResultSet rs = stmt.executeQuery(sql);

        int i = 0;
        while(rs.next()) {
            int traffic=rs.getInt("교통량");
            double avgTem=rs.getDouble("평균온도");
            // 월강수량-교통량그래프로나타내기
            // int traffic=rs.getInt("교통량");
            // double monPre=rs.getDouble("월강수량");
        %>
        arr.push(['<%=avgTem%>', '<%=traffic%>', 'gold']);
        // arr.push(['<%=monPre%>', '<%=traffic%>', 'gold']);
    <%
        }

        if(rs != null) rs.close();
        if(stmt != null)stmt.close();
        if(conn != null)conn.close();

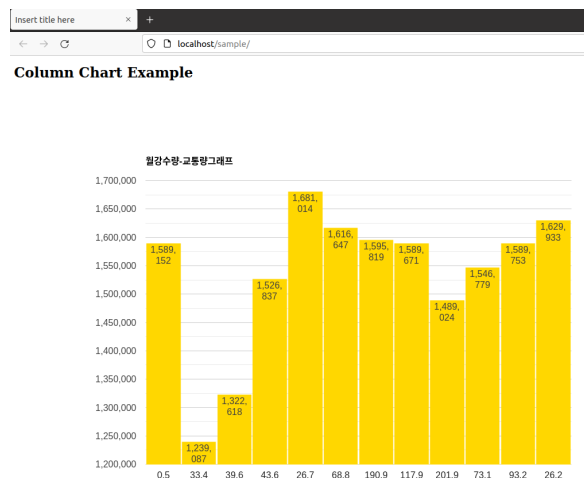
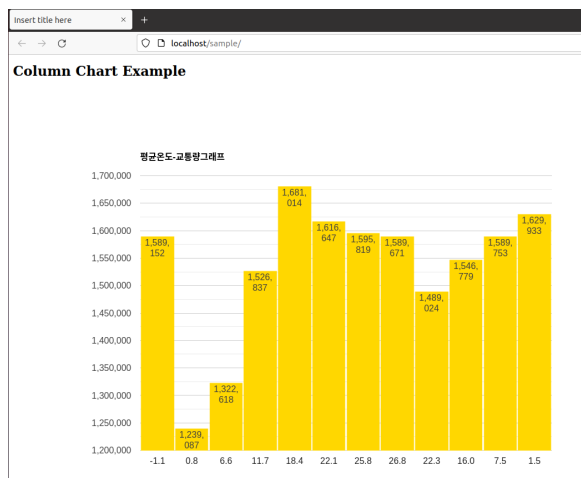
        conn.close();

    }catch(Exception e){
        out.println(e);
    }
    %>
    console.log(arr);
    // Create the data table.
    var data = google.visualization.arrayToDataTable(arr);
    var view = new google.visualization.DataView(data);
    view.setColumns([0, 1,
        { calc: "stringify",
          sourceColumn: 1,
          type: "string",
          role: "annotation" },
        2]);

    var options = {
        title: "평균온도-교통량그래프",
        // title: "월강수량-교통량그래프",
        width: 1200,
        height: 800,
        bar: {groupWidth: "95%"},
        legend: { position: "none" },
    };

    // Set chart options
    var chart = new google.visualization.ColumnChart(document.getElementById("columnchart_values"));
    chart.draw(view, options);
    }
```

```
</script>
</body>
</html>
```

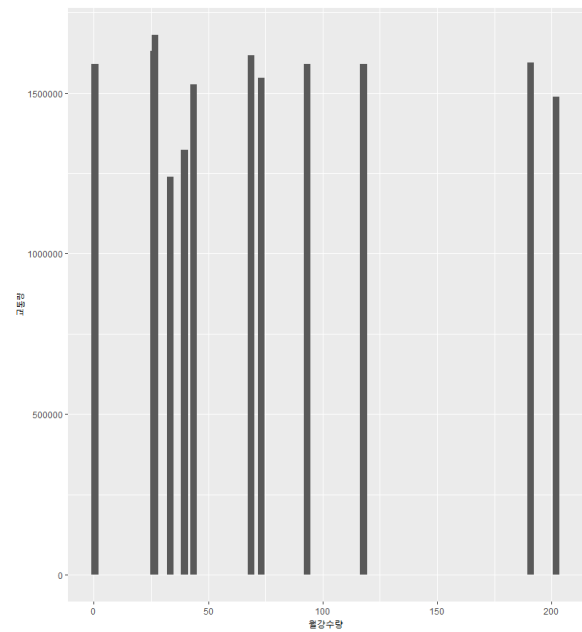
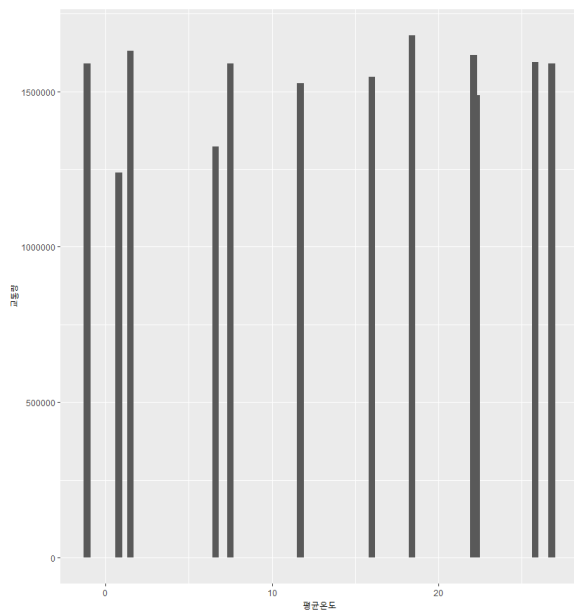


• R을 이용한 시각화

```
# 시각화를 위한 패키지 ggplot2
library(ggplot2)

# data에 쿼리 결과 저장
data <- dbGetQuery(conn, "SELECT TO_CHAR(T.MDATE, 'YYYY/MM/DD') AS 년월
, T.REGION AS 지역
, TRA.지점
, TRA.교통량
, T.AVGTEMPERATURE AS 평균온도
, P.MONPRECIPITATION AS 월강수량
FROM TEMPERATURE T
INNER JOIN PRECIPITATION P
ON T.MDATE = P.MDATE
LEFT OUTER JOIN (SELECT TO_CHAR(COLDATE, 'YY/MM') AS 년월
, COLPOINT AS 지점
, SUM(VOLTRAFFIC) AS 교통량
FROM TRAFFIC
GROUP BY TO_CHAR(COLDATE, 'YY/MM'), COLPOINT
ORDER BY TO_CHAR(COLDATE, 'YY/MM') ASC, COLPOINT DESC) TRA
ON TO_CHAR(P.MDATE, 'YY/MM') = TRA.년월
WHERE TRA.지점 LIKE '경수%' # 경수대로 데이터만 출력
ORDER BY TRA.지점 ASC")

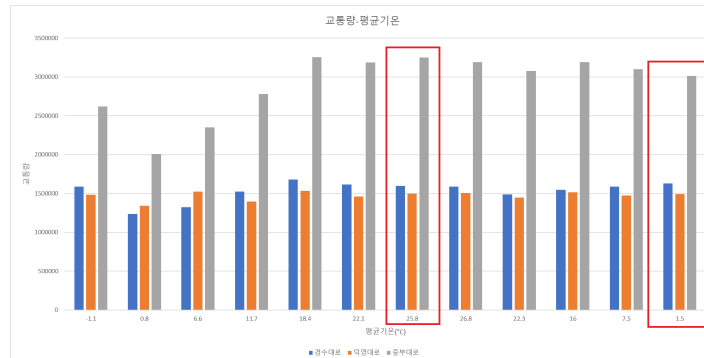
# 평균온도 - 교통량 막대그래프
ggplot(data, aes(x=평균온도, y=교통량))+geom_bar(stat="identity", width = .4)
# 월강수량 - 교통량 막대그래프
ggplot(data, aes(x=월강수량, y=교통량))+geom_bar(stat="identity", width = 3)
```



결론 및 의견

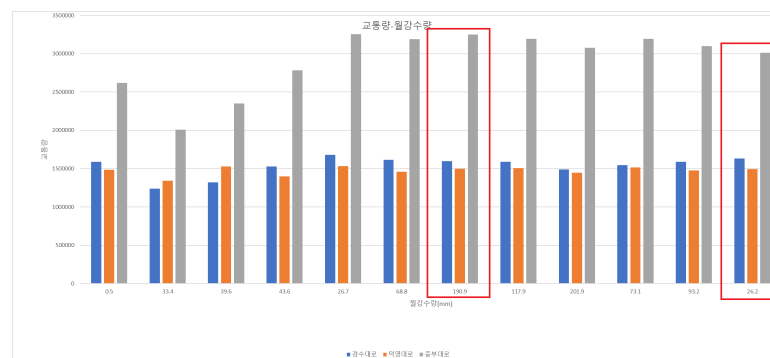
• 교통량과 평균기온의 상관관계

- 아래의 그림에 표시된 부분을 보면 평균 기온이 24도 차이가 남에도 불구하고 교통량은 유의미한 차이를 보여주지 못한다.



• 교통량과 강수량의 상관관계

- 아래의 그림을 보면 강수량이 차이가 큰 여름과 겨울을 비교했음에도 불구하고 교통량에는 크게 유의미한 차이를 보여주지 못한다.



• 결론 및 의견

- 위의 두 데이터를 보면 평균기온과 강수량은 교통량과 크게 상관이 없음을 알 수 있다. 시각화한 데이터를 보면 교통량은 겨울이 될수록 줄어듦과 봄부터 점차 증가함을 볼 수 있다.
- 이와 같은 데이터를 통하여 가설을 두 가지 생각해 볼 수 있는데, 첫 번째 가설은 여행이 비교적 적은 겨울철에는 교통량이 적고 날씨가 풀리는 4월부터는 점차 여행을 떠나는 인구가 많아져 교통량에 변화를 줄 수 있다. 두 번째 가설은 1~3월의 경우 눈이 내리기 때문에 교통량이 줄어 들 수 있다.