

## Preparing Data

1. Write a DATA step that reads the pg1.storm\_summary table and creates an output table named Storm\_cat5.

```
data storm_cat5;
  set pg1.storm_summary;
run;
```

2. Include only Category 5 storms (**MaxWindMPH** greater than or equal to 156) with **StartDate** on or after 01JAN2000.

```
data storm_cat5;
  set pg1.storm_summary;
  where StartDate>="01jan2000"d and MaxWindMPH>=156;
run;
```

3. Add a statement to include the following columns in the output data: **Season**, **Basin**, **Name**, **Type**, and **MaxWindMPH**.

```
data storm_cat5;
  set pg1.storm_summary;
  where StartDate>="01jan2000"d and MaxWindMPH>=156;
  keep Season Basin Name Type MaxWindMPH;
run;
```

4. Add an assignment statement to create **StormLength** that represents the number of days between **StartDate** and **EndDate**.

```
data storm_length;
  set pg1.storm_summary;
  drop Hem_EW Hem_NS lat lon;
  StormLength = EndDate-StartDate;
run;
```

5. Open the **pg1.storm\_range** table and examine the columns. Notice that there are four wind measurements for each storm.

- I. Create a new column named **WindAvg** that is the mean of **wind1**, **wind2**, **wind3**, and **wind4**.
- II. Create a new column **WindRange** that is the range of **wind1**, **wind2**, **wind3**, **wind4**.

```
data storm_windavg;
  set pg1.storm_range;
  WindAvg=mean(wind1, wind2, wind3, wind4);
  WindRange=range(of wind1-wind4); run;
```

6. Add a WHERE statement that uses the SUBSTR function to include rows where the second letter of **Basin** is *P* (Pacific ocean storms).

```
data pacific;  
  set pg1.storm_summary;  
  drop type Hem_EW Hem_NS MinPressure Lat Lon;  
  where substr(Basin,2,1)="P";  
run;
```